The Thesis committee for Sarah Lynn Pfannenschmidt

Certifies that this is the approved version of the following thesis:

‘How do we evaluate this?’:

Perspectives on evaluation criteria for digital scholarship from the digital humanities community

APPROVED BY

SUPERVISING COMMITTEE:

Supervisor: _____

Tanya Clement

_____

Patricia Galloway

‘How do we evaluate this?’:

Perspectives on evaluation criteria for digital scholarship from the digital humanities community

by

Sarah Lynn Pfannenschmidt, BA, MA

Thesis

Presented to the Faculty of the Graduate School

of the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Information Studies

The University of Texas at Austin

May 2013

ACKNOWLEDGEMENTS

The author expresses her gratitude to her thesis committee, Dr. Tanya Clement and Dr. Patricia Galloway, for their excellent feedback and suggestions in the development of this work.

Thanks also to Dr. Lecia Barker for her helpful comments.

The author acknowledges that this study would not have been possible without the generous cooperation of the study participants. Thank you for sharing your time and perspectives on this topic.

Last but never least, the author recognizes the patience, incisiveness, and good humor of Ms. Michelle Seymour. Thank you for faithfully pointing back to who is able to accomplish any of this.

‘How do we evaluate this?’:

Perspectives on evaluation criteria for digital scholarship from the digital humanities community

by

Sarah L. Pfannenschmidt, (MSIS)

The University of Texas at Austin, 2013

SUPERVISOR: Tanya Clement

Since the advent of the World Wide Web, there has been an increasing influx of digital scholarship. Such scholarship is not always recognized as legitimate, in part because digital work is still in its 'incunabula phase' and also because the staggering variety in tools, user communities, etc. engenders a host of potentially competing evaluation priorities. These concerns have created a pressing need for appropriate evaluation criteria to fairly assess digital projects. Though this topic has received substantial attention in the scholarly literature, discrete solutions and the establishment of firm yet flexible evaluation criteria remain elusive. This paper presents a pilot study that sought to clarify the following: what criteria participants use to evaluate digital scholarship, the place of digital tools in the evaluation of scholarship, who should evaluate digital projects, the role of stated intentions in the formation of evaluation criteria, what role the TEI might play in evaluation of text encoding, and finally how this role would be practically implemented. The study indicated that despite the complex nature of the topic, a number of practical solutions may aid in the legitimization of digital scholarship. In particular, including

a statement of intent that explains the methodology of the project goes a long way in establishing the relationship between the content and the tools and the criteria to evaluate both components. Two potential roles for the TEI community also emerged: (1) to provide counsel and formative assistance with ongoing projects in a manner targeted towards project evaluations and (2) to consider including dedicated reviews section in the Journal of the Text Encoding Initiative to feature project evaluations and accept submissions for review. This publication is an ideal online platform for the discussion of review guidelines and may help to clarify what evaluation criteria are necessary to promote fair and accurate assessments of digital projects. Determining what to evaluate and how to do so are perennially relevant questions, and as digital scholarship continues to develop, it will become more important than ever to develop a better of understanding of what we value and why we value it.

TABLE OF CONTENTS

Part I: Introduction

With the advent of the World Wide Web, the creation, publication, and dissemination of digital materials has experienced exponential growth. As computers become a central component in the research process, scholars are increasingly faced with a field that is using digital publications, online forums, and social media. Budget cutbacks in both libraries and the publishing industry has led to an increase in digital scholarly publications as well as online copies of print journals [1], leading to increasing access to the research of our peers. Digital media has also made it possible to interpret and visualize large data sets in new ways [2], even as the globalization of the web has initiated an increasing call for open access to research. However, as research moves into the digital age, there has been more discussion about precisely what is meant by the term 'digital scholarship' [3, 4]. As Laura Mandell explained, digital scholarship is more than "simply scholarship that takes place in digital media" [5]. Digital scholarship encompasses something more: the curation, collaboration, and exploration of what is possible using digital media, presented and framed in a rigorous methodological way. When all of these new considerations are added into the pot of evaluation of research, it may appear difficult if not impossible to judge these materials through the lens of traditional scholarship. Thus, the definition of what we mean by scholarship has far reaching implications for what work is accepted, valued, and rewarded by the academy and other evaluative bodies.

The trouble with defining digital scholarship is nowhere more evident than in the field of humanities computing, also known as digital humanities (DH). There have been several attempts to clarify what the digital humanities are, what they do, and their place in the academy [6-8]. One definition of the field is that it is "the thoughtful use of computing in humanistic inquiry and the thinking through of computing from the perspective of the traditions of the humanities" [9]. This definition strikes at the dualist nature of digital humanities: that computing and the

humanities each bring their specific priorities and methodologies together to meet 'in the middle' via research. The application of computers to the study of the humanities is not new; projects in this area have existed since at least the 1960s. However, the field as a whole is still working through what has referred to as its digital incunabula phase [10]. By using the term 'digital incunabula', the implication is that scholarship is still deciding on what is expected of projects, both in terms of tools and methodology, in order to be considered credible. Establishing a clear picture of what can be done in a project, both in terms of determining what to do [11] and how to do it [12], is still difficult. However, the outgrowth of digital humanities centers and the sheer amount of literature published on related topics demonstrates that the field of humanities computing is here to stay [13]. As such, there is a pressing need for the community to determine how to evaluate digital projects. In some ways this problem illustrates a feedback loop because the definition of scholarship is an expression and application of evaluation criteria. Frankly, we understand what digital scholarship is based (in part) on its evaluation.

The history of the discussion on the evaluation of digital scholarship is long and complex. Multiple papers, conferences, and other deliverables have been released in the last ten years in an attempt to address this topic. For example, the six essays in the 2011 *Profession* discussed a number of salient points, including the fact that digital scholarship needs to be reviewed by experts in both the content and the technology used in a project [14]. Interest in the topic has also prompted the reissue of many key discussions. In a recent issue of *Journal of Digital Humanities* (hereafter JDH), several seminal works on evaluating digital scholarship have been aggregated to allow the community to see what has been discussed and what still needs to be addressed [15]. Most recently, NINES has conducted a series of summer sessions with a focus on this topic, and is in the process of producing a whitepaper of evaluation suggestions and guidelines for digital projects. Three key suggestions emerged from these sessions, namely that evaluators 1) "review and assess the project in which the medium in which it was created, 2) recognize the intrinsically

collaborative nature of digital projects, and 3) consults specialists in relevant disciplines regarding the various components of the project" [16]. Though the dialogue is ongoing, a number of stumbling blocks have remained, not least of which is trying to decide how to implement suggested criteria consistently and fairly in a field that is famously (or infamously) diverse and where no single body has authority to enforce best practices.

While it is impossible to cover the entire history of the question here, there are a number of important points that are a matter of considerable debate. First is the consideration of how digital technology creates "new approaches to knowledge construction" and how to evaluate this in a digital project [17]. One of the biggest concerns from the DH community is that reviewers have applied traditional print-based evaluation criteria to digital work, in part because the humanities are accustomed to reviewing print materials, such as book reviews of monographs or other publications. This type of review is designed to travel a clear route through specific assessments and end with a 'stamp of approval' demonstrating the scholarship of the material. Supporters of print-based assessments have advocated for the use of similar criteria in the assessment of digital projects, with the implication that computers do/do not change some of the practical considerations for editing and publishing texts [18]. However, other scholars have demonstrated that digital materials add their own unique ontological and epistemological perspectives to the discussion [19] through their very 'digital nature'. As such, there is a concern that digital projects are neither accurately nor fairly evaluated by applying traditional criteria [20, 21].

The second and an especially contentious point in the application of print-based criteria to digital projects is how digital tools are evaluated. Digital scholarship uses technology to look at material in new ways and to produce new knowledge. Many of these tools are built or adapted for a project in order to help the project achieve its goals. As such, digital scholarship does not always "'look' like traditional academic scholarship" even though it still requires an examination of

"how the work contributes to and advances the state of knowledge of a given field or fields" [22]. At present, digital tools do not have clear, articulated, and widely accepted guidelines for evaluation within the humanities. This is partly due to the fact that it is often unclear what aspect of the tool is being evaluated: its usability, its theoretical framework, both, or something else entirely? [23]. Evaluating digital tools has also not necessarily been the purview of the humanities; typically evaluation of the technology from a tool-builder perspective has been relegated to computer science, information studies, and usability. Clarification within the community of what criteria are needed to assess digital tools may develop given time and further discussion, but trying to define evaluation criteria for tools in the interim is a delicate task. This delicacy is in part due to the fact that criteria are double-edged weapons; while they may provide structure and clarity to fairly assess digital tools, they may also close off other avenues or become limiters to innovation if they are not carefully constructed so as to be flexible and widely applicable [24].

Another complication in the assessment of digital projects is that there are multiple types of evaluation and the criteria change depending on both the project and evaluative goals. Different reviews have different purposes and priorities and therefore require corresponding evaluation criteria. For example, a grant review may account for the potential contributions of the proposal and the value for money investment, while journal reviews look for methodological rigor and concrete outcomes that add to a discussion. Perhaps the most troublesome evaluation in relation to assessment criteria is for tenure and promotion. Convincing non-digital colleagues of the scholarly merit of digital projects is challenging since the effort that goes into digital projects is not always easily quantifiable or demonstrable (at least in traditional ways) [25, 26]. Digital projects frequently involve collaborative and interdisciplinary work, which is not only in stark contrast to the established single author monograph, but also makes it difficult to distinguish the contributions in a group project [27]. Evaluation for these cases depends partially on the project's disciplinary field and partially on the values of a specific institutional review board.

It can therefore be difficult to understand the expectations of the evaluation or uniformly apply criteria across institutions.

However, it has been noted that a project's value may lie in revolutionizing the way we think rather than in successfully meeting all of the evaluation criteria. Many digital projects push theoretical and technical boundaries, but their value to the community may be in their failure [28, 29]. Failure, like success, adds to the community dialogue about what works (or does not work) and why. As such, projects may impact a field in ways that evaluation criteria do not naturally take into account, in part because the criteria are geared towards prizing success as the most desirable outcome. The best experiments are the ones that are equally informative, whether they succeed or fail. Therefore, it is important to identify and explicitly state what criteria are being applied when a digital project is evaluated, since much of the criteria currently in use are implicit and do not account for this productive failure. That is, the community takes for granted that 'success' is evident in project and that the evaluator will know what to look for and how to reward it.

Identifying who is the best person or group to assess digital projects is closely tied to determining what aspects of a project should be evaluated. In order to ensure a fair review, the quality of evaluations depends on identifying and including the appropriate evaluators. In other words, evaluators presumably are able to evaluate because they demonstrate a certain level of expertise in a specific area. However, it is not always clear what criteria determine the expertise of a reviewer, or who identifies the successful attainment of this expertise. Such a concern is especially problematic in relation to the evaluation of digital tools. Significant advances have been made in this area over the last ten years with the creation of peer review groups like NINES and 18thConnect (and soon MESA). Scholarly organizations including the MLA [30], the AHA [31], as well as institutional bodies [32] and grant bodies [33] have also been exploring the type of reviewer that is best suited to evaluating digital work. Many of these groups have published reports encouraging review committees to evaluate materials by

engaging with outside experts who can vet specific aspects of a digital project. However, these reports are limited in two ways. First, the bodies can only encourage the community to best practices; they cannot force them. It is also a limitation that many of these bodies are focused on a specific area within the humanities rather than considering how evaluation criteria differ between field specialties. Therefore, a linguist assessing the digital work of a historian may have a different set of priorities for a project, and vice versa, that impact the criteria applied in the evaluation. Different field-specific criteria may lead to interesting dialogues, but it can also serve to shut out some of the aspects of a project that are central in the project's methodology.

Because evaluations do not take place in a vacuum, it is necessary to make reviews available where appropriate to share results and add to knowledge of what criteria are being applied in individual contexts. Though naturally there are evaluation criteria that differ among fields, there may also be a number of similarities. Making these similarities explicit edifies the wider community in determining what evaluation criteria are important in terms of a hierarchy of value. Journals have traditionally been a viable forum for the presentation and discussion of evaluations for print scholarship and are now also prolific arenas for digital communication [34]. Yet though there are multiple DH focused journals that feature reviews of books published on DH topics, there is a lack of featured project reviews for current scholarship. Some progress has been made in aggregating lists of projects that have undergone some form of peer review, as demonstrated by NINES. However, while NINES provides assessment guidelines and incorporates reviewed projects into its interface via Collex, the documentation of the content of specific reviews is not necessarily available [35]. While this may reflect privacy precautions, it is notable that the pioneering peer-review group in the field of digital humanities research does not include these materials with the projects on the site.

This paper describes a study that considered evaluation criteria for digital projects as shared by members of the DH community. In particular, the study

considers how digital projects, especially digital scholarly editions, are currently evaluated, and what role the Text Encoding Initiative (hereafter TEI) community may have in this evaluative process. The TEI is in a unique position as both a standard for best practices and as developing set of tagging guidelines that have been evolving for three decades. The breadth and depth of its community members encompasses users of all levels, skills, and humanities disciplines. As such, this study asked the DH community to identify any evaluative contexts in which the TEI might be involved (and how this would relate to the type of evaluation being undertaken). Finally, the research findings suggest that there are a number of practical steps in the identification of evaluation criteria for scholarly work.

There are two caveats to this study on how participants were asked to comment on some of the problems encountered between traditional and digital scholarship [36]. When we say 'digital projects', this paper is predominantly referring to digital editions because these are the materials that participants used as their examples. However, digital editions are an interesting starting point for the application of evaluation criteria because they blend editorial theory and practice with "the combined strengths of database and narrative" [37]. Thanks to the digital environment, editions can include more materials and present these objects in multiple ways that would be otherwise impossible in a codex (for example, interactive visualizations). Through technology, digital editions push the theoretical boundaries about the text as well as our understanding about what a text or an editor or a reader means in a digital environment. Many scholars have noted that because of these enormous possibilities, the digital environment presents a number of practical and theoretical challenges not only to the editor [38-41], but also to the evaluator. However, defining evaluation criteria is challenging because digital scholarly editions have multiple forums where they must demonstrate their scholarly value. In particular, digital scholarly editions are frequently part of evaluations for tenure or promotion [42]. There are two concerns in particular that further complicate matters. In general, editions have been afforded a diminished

value in some fields where textual criticism is seen as 'secondary' scholarship [43]. Also, many traditional departments are unsure how to assess digital materials that confound print-based criteria in part or in whole.

The second half of the paper focuses on a discussion about text encoding and its place in evaluations of digital scholarship. The reasoning behind this focus is that many digital editions use text encoding to represent their materials online, and a significant amount of these utilize the Text Encoding Initiative's (hereafter TEI) Guidelines. Multiple grant bodies (such as the National Endowment for the Humanities) have recommended these guidelines as the accepted standard for encoding texts in digital scholarship. The fact that the guidelines are in their fifth permutation demonstrates that interest in the TEI has been rich and longstanding in relation to digital projects [44]. The TEI is not only a best practices standard, it is also a theoretical approach to the markup of text [45-47]. As such, the TEI functions as an interesting representative in the consideration of how digital and traditional editorial practices work together and how the scholarship of these projects are evaluated in relation to their text encoding practices. It was not within the purview of this paper to determine if TEI tags are sufficient for all editorial needs; studies in this line have been done elsewhere with fascinating results [48]. Instead, the focus was on the application of the TEI tags as part of evaluated digital projects, as well as on the TEI Consortium and community's role in the larger question of how to assess these materials in terms of their 'scholarship.'

Part II: METHODS

Study Design:

The original project design utilized a survey-based approach in order to gain a broad view of how the community defines criteria for evaluating digital projects as well as what identifiable standards are currently in use. Short interviews were planned to follow the completion of the initial survey using a more focused set of questions derived from the survey's key findings. Participation in the initial survey was limited. However, the thoughtfulness of the comments provided in response to the short answer questions indicated that there was significant interest in engaging with the topic at length which was not effectively facilitated by the survey format. Thus the original survey design was changed in favor of an interview approach that would produce more detailed and comprehensive information on the topic.

Study Participants:

The intended audience for both the survey and the interviews was the general digital humanities community, especially those who had project experience with critical editions of texts and text encoding. No criteria for participation were set beyond an interest in hearing from individuals who associated themselves with the TEI and/or the digital humanities community. In particular, no specifications were made requiring that participants be members of the traditional academic community.

The initial Call for Participation for the survey was sent to three email listservs: The Humanist Discussion Group (hereafter Humanist) hosted by King's College, London; the TEI listserv, out of Brown University; and the Digital Medievalist (DM), hosted by the University of Lethbridge (see Appendix 1 for the survey CFP). The rationale for requesting participants from these listservs was that these communities are well characterized. For instance, all three of these listservs

have large and active communities; there are 744 members on the subscriber list for DM as of May 2012, 2,149 active memberships for Humanist, [1] and 1,033 subscribers to the TEI listserv.[2] All of these listservs are moderated, so the participant request had to be approved for submission and distribution among list members. The listservs are not restricted to a specific academic discipline. Humanist allows for any persons interested in humanities computing and the digital humanities to subscribe to the list. Individuals are welcome to join the DM listserv if they are associated with or interested in the study of the medieval period and digital media. Finally, the TEI listserv permits anyone interested in or working with TEI to join the list. Because the survey questions did not focus on any single discipline and the larger issues addressed in the survey are of relevance to the entire community, the principle researcher felt that asking across traditional disciplinary boundaries would permit a greater number of people to contribute to the discussion.

Interview participants were solicited from three sources. One group of interview participants was contacted through the TEI Conference attendance list. The conference was held at Texas A&M, College Station, TX, on November 7-10, 2012. The invitation to participate was sent to all registered conference attendees via the conference email list. This CFP was sent with permission from the conference organization committee and was personally addressed to each participant (see Appendix 2). The researcher was given a table in the conference registration and general meeting area and was available throughout the day of November 9th for both pre-arranged and walk-in interviews. Any participants that

[1] The numbers provided here are figures designed to give a general sense of the size of each community. Membership or subscription to these listservs should not be counted solely in terms of individual users. In the case of Humanist, for example, membership can also be extended to groups and redistribution points. It is the same for DM and the TEI listservs. The numbers given here, therefore, likely under and over represent the actual listserv readership. Also, these numbers cannot indicate the number of persons who actually participate on the listserv by posting, etc.

[2] As stated on the TEI website, "Anyone can subscribe to the list…and everyone is welcome." Being on the TEI listserv, however, does not automatically confer membership in the TEI Consortium, which is a paid membership (the listserv is free). The Call for Participation sent to the TEI community was sent to the general community of individuals who are interested in or use TEI, regardless of their status as official members or subscribers of the TEI Consortium.

expressed interest in being interviewed but were unable to attend the conference or to be interviewed in person at the meeting were contacted afterwards to arrange for a remote interview via Skype. Other participants were solicited from the original CFP for the survey; these individuals contacted the researcher and indicated a willingness to speak further on the survey materials. The final group of interview participants was personally invited to speak with the researcher via an invitation from the principal researcher's faculty sponsor, who is a member of both the digital humanities and the TEI communities.

Participant demographics for the interviews: five female and fifteen male participants were interviewed. Ten of the twenty participants held tenure track positions at research and teaching institutions across North America and Europe. Of these individuals, at least five were tenured professors, and at the time of the interview a sixth participant was undergoing tenure review. The remaining ten participants included individuals who were employed in what has been referred to elsewhere as 'alternative academic' positions: independent scholars, librarians, adjunct faculty, etc. [49]. All of the participants had experience in the creation of at least one digital edition of a text or texts. It is important to note that because I was interested in asking the community about the use and evaluation of the TEI in digital projects, the participant pool was predominantly comprised of individuals who use or have used the TEI and have positive opinions about its use as a standard for text encoding. Only one participant out of the twenty openly indicated that they did not value the TEI as a relevant or useful community standard.

Survey:

The survey (see Appendix 3) consisted of 20 questions designed to solicit participant opinions on the following topics: current methods of evaluation for digital projects; the sufficiency of these methods in the fair evaluation of scholarship; the perceived place of text encoding in the evaluation process; the present and future role of the TEI Consortium in the evaluation of its own standards as applied to digital projects; how this evaluation would be undertaken, and by

whom. The survey was hosted on Qualtrics, a marketing-research platform with web-based data collection and analysis services. Questions were written by the principle researcher in consultation with the faculty sponsor and were approved prior to the survey launch. The survey was drafted, tested, and made available to the target community from November 5, 2012 until January 21, 2013. A total of 26 people started the survey with a maximum of 9 responses per question. However, the Qualtrics service did not track individual users. Therefore, it was unclear which participants of the original 26 answered each question or completed the entire survey. Participation in and responses to the survey were both anonymous.

Interviews:

Interviews were conducted with a total of 20 participants. Of these 20 interviews, 13 originated from the TEI Conference attendee list, 1 was a follow-up from the initial survey, and 6 were by personal invitation. All participants who began the interview process completed the interview. A preliminary interest in being interviewed was expressed by at least 33 participants; due to numerous factors, 13 participants were lost to follow up and were unable to be interviewed.

Interview questions were adapted thematically from the survey questions (see Appendix 3). Interviewees were allowed to opt out of any questions and also to determine the length of the interview based on their availability. Individual interviews lasted from 10 minutes to over an hour long. The principle researcher and interviewer requested participants' permission to record the interview session in order to facilitate accuracy in the data and to maintain a conversational style. Professor Lecia Barker, research associate professor at the School of Information at the University of Texas at Austin, provided consultation and feedback in the development of the consent to record document. This form (see Appendix 4) was provided in person or sent via email to participants when they arranged for a day and time to be interviewed, and consent was given to record the interview with two exceptions. Notes were taken by hand during these two interviews. For those who

agreed to the consent form, their interview was recorded using a single portable recorder. There were 10.5 hours of total recorded interview time. Interviews were conducted between November 9, 2012-March 8 2013. Nine of these interviews were conducted in person; the remaining interviews were conducted remotely.

Data Analysis:

Survey responses were anonymous and compiled into a summary report by Qualtrics. The small sample size excludes the need for any descriptive statistical analysis.

The recorded interviews were uploaded as mp3 files to a private, password protected computer and transcribed into text files. Transcribed files totaled 151 pages and were coded and scrubbed of personal or identifying information in order to ensure the anonymity of the participants. Additionally, each file received an alphabetical designation as the file name. The transcripts were reviewed, and data was extracted and compiled by topic. Audio files were deleted following a final review of the transcripts. The general trends and opinions as well as the specific responses from the interviews will be discussed in Part III: Results and Discussion.

Part III: RESULTS AND DISCUSSION

A: What criteria do participants apply in the evaluation of digital projects?

During the interviews, the principle researcher asked six topically related series of questions. The first series was intended to identify two things: what aspects of a digital project determine the 'success' or value of the project, and what criteria individuals used in order to make these assessments. The interviews frequently began by inviting participants to give examples of projects that they felt were indicative of 'good', 'scholarly', 'successful', or 'exemplary' digital work. The point of using these ambiguous terms was to see how participants defined these words in relation to evaluation criteria. The follow up questions asked participants to explain *why* they identified the cited projects as 'successful' and how the projects demonstrated this success. The way participants described specific example projects further clarified the criteria that they applied in the assessment of digital work. Participants were also invited to explain which evaluation criteria were more or less important in the overall assessment process in order to determine which of these criteria were most important.

Overall, there was no simple answer given in response to these questions. Participants' responses were extensive and varied, and highlighted the complex nature of defining evaluation criteria. The general trend of the responses was to express evaluation of a digital project in terms of both the technical aspects, such as usability, accessibility, interface design, innovation of tool use, etc., and the importance and impact of the scholarly content for a given field or fields. The majority of participants agreed that the technical aspects of a project played a role in evaluation, though the extent of this role was determined by the individual priorities of each participant. Thus, all projects cannot be evaluated on the same 'technical level'. While the majority of participants ranked the evaluation of technology secondary to content, a small number of participants suggested that assessment must consider all aspects of a project in order to perform a fair assessment of a project's value.

One word that frequently appeared in participants' selection and descriptions of good digital projects was the term 'useful'. Participants defined usefulness in a digital project in a number ways, and associated different evaluation criteria with how they defined this term. One group of participants focused on engagement with a target community as the primary indicator of whether or not a project was useful. For example, Participant A explained that, "A good project is a project that is *useful for a community of reference*, maybe more than one community of reference..." (emphasis mine). In Participant A's statement, there is a cyclical implication that evaluation criteria for a project depends on the priorities and values of a specific existing community. This "community of reference" decides what is useful in a project based on the community's needs. It would therefore be of central important for a project to identify its target community (or communities) and address the user needs of this group. On the one hand, this perspective is a logical corollary to the view that those who use a project are also its most appropriate evaluators. However, defining a 'community of reference' for a digital project is not always straightforward. The community of users may be undefined or the project may need to include multiple user groups. From Participant A's response, it is unclear who determines the community's hierarchy of needs that the project will ideally address. While identifying the needs of a community and the implications for the evaluation of digital scholarship is itself an important topic, in the end this statement indicates that 'usefulness' can be tied to to a project's ability to meet the needs of a target audience.

Responses from other participants suggested that a project's usefulness was not dependent solely on meeting the needs of the target community, but rather how well that project was able to reach people outside of that community. One participant said that they evaluate the usefulness of a project by inquiring, "Is [the project] useful or interesting to the non-expert?"[Participant G]. The main implication of this statement is that broad relevance or transferability is more useful than conforming to a narrow or deep niche. In other words, the project must not

only be useful to one or two highly specialized people, but also to a wider group of potential users. However, it is unclear how a project effectively demonstrates that it has considered the needs of the "non-expert." Indeed, the term "non-expert" in this statement is ambiguous; it does not indicate whether such a user is unfamiliar with the project's content, tools, or both, or even if it refers to something else entirely.

Other participants defined usefulness apart from existing or potential user communities and focused instead on a project's pedagogical value. Participant R explained that her digital project's "emphasis is *more on the learning outcome* …I think [my colleagues] want scholarly activity as well, but [the pedagogical usefulness of the project] was kind of the *first* evaluation." The first priority of Participant R's colleagues was to consider how her project was used in the classroom. By this definition, usefulness is directly tied to the ability of a project to be used as a teaching tool. It is no surprise that the way in which a given community defines and applies their specific interpretation of 'usefulness' will reveal the priorities of that community. The perspectives expressed by Participant A, G, and R share an emphasis on *to whom* a project was useful rather than *how* it was useful to a community of users. However, the variation between the communities to which each participant gave priority illustrates the challenges of this approach to defining usefulness. Priorities among communities differ by necessity, and it is unclear how evaluation criteria could or should reconcile these interests.

Participants also discussed an entirely different definition of 'useful' that focused on the cohesiveness and presentation of the project. The cleaner and more seamless the technical aspects of a project, the more positively participants evaluated a project. Several participants included comments on usability as part of their assessment of a project's aesthetics. The rationale for these opinions appears to be connected to the influence of the overall design of the interface and use of certain tools in providing access for the project content. For example, Participant K explained that part of evaluation is dependent on "the way in which the content is presented." In K's opinion, if the materials are presented in a "particularly

innovative" way, then there is added value to the project, though it is unclear what constitutes the largely subjective designation of 'innovative'. Participant G explained that "the things I want to see in an interface typically are mostly ease of getting to the information that the user wants…a search interface that's usable…that lets the user primarily read the document, [that is] pretty to read and…present[s] a user with the capabilities" to read the work in multiple ways. In G's response, the "ease" of access determines how useful and therefore how valuable a project is.

Yet aesthetics are often the most fragile aspects of digital projects, as well as the most subjective to evaluate, especially in terms of usefulness. Participant L indicated that the presentation of the material "just has to be basically understandable. So we're wiling to accept a certain amount of clunkiness if there is a good reason for it." Participant L's response suggests that usability *and* aesthetics are less important in determining a project's value than the provision of access. There is an inherent tension expressed here between ease of use and the vulnerability of the technology. As tools and platforms change or update, there is a potential for problems with corrupted data, broken tools, etc. Participant J voiced this concern and additionally noted that, "if you have a wonderful project, but it's hard to navigate or it's hard to visualize depending on your device, or [it is hard] just trying to find it, trying to access it itself… I think that would take away from the scholarly worth of the project." As such, it seems that the 'usefulness' of a project and demonstrated 'scholarliness' is to some extent tied to its aesthetic presentation and usability. If the project is not 'useful' in terms of access (users cannot find the data, navigate the interface, or use the tools, etc), then the project does not have the opportunity demonstrate its full scholarly merit. Ultimately, access may be the best word to describe this aspect of 'usefulness', because it involves more than simply the design aesthetics of a project. From Participant J's response, the issue appears to be whether or not the scholarship has the chance to demonstrate its real worth scholarly or otherwise, and whether or not the design and the tools used in the project are aiding or hindering in this endeavor.

Another participant definition of 'usefulness' involved the long-term sustainability of the software, platforms, and tools used by a project in order to produce and share new data. Notably, participants who gave this opinion were also tool builders or were responsible for the long-term preservation and maintenance of digital projects. In order to continually impact the scholarly field of interest, projects need to be maintained and kept available for future work. If the project cannot be sustained, then its impact on scholarship may be diminished or at least shortened. As Participant J said, if "no library can hold onto [the project] or you can't preserve it yourself that should factor into, again, the scholarly merit of it. Is it something that *will continue to contribute*?" (emphasis mine). Participant D also connected the production of new data over a period of time to the scholarly impact on both the subject area and the target community:

> I try and evaluate…how useful [the project] is…is it freely and openly available and is it openly licensed…does it not only give you a shiny website where you can search by whatever criteria that make sense for that particular resource, but does it also give you access to the underlying data…[so that] you can do other things with it that [the project creators] might not have predicted?

In some ways, interest in the production of raw data and concern for the long-term sustainability of the project between evaluating a project based on who uses the project and the project's aesthetics. In J and D's opinions, the future value of a project plays a role in determining its immediate usefulness. If there is a potential to continue using the project's data long after the original materials are initially made available, then the community will continue to benefit from it. As such, the aesthetics may not matter as much *in the long term* if the raw data can be used to continually produce new scholarship.

One interesting result of participant definitions of the term 'useful' was how the relationship between tools and content was described. Overall, participants devoted more time to considering the technical aspects of a project and the impact on the project's value. However, though content was discussed more briefly, it was

also given primacy of place in the evaluation of the global 'success' of a project. Many of the participants who had described *technical* aspects of a project when asked how to evaluate a digital project (or a project's technical outcomes, such as usability, interface, improvement of other digital tools, etc.), concluded that usefulness of *content* was the deciding factor of whether or not the material was worthy of both scholarly attention and evaluation. One participant explained that there are different priorities for levels of evaluation, with content coming first:

> I think still the first priority is the content presented and synthesized to be of scholarly merit. So, similar to a paper, you know, are they making the right arguments? Are the arguments presented well? And I think the same thing happens in a digital project, and then on top of that, there is the *added element* of interactivity and usability [Participant J] (emphasis mine).

There was an implicitly understood condition in this statement that 'good' content acted as the 'trump card' in evaluation because it was the demonstrable scholarship of the project. Despite this implication, good content was described by participants in the least amount of detail and with vague reference to unarticulated but apparently common knowledge about what constitutes scholarship. The underlying assumption is that scholarship is self-evident and is primarily manifested in digital projects as the content.

A small number of participants (15%) explicitly pointed out that they did not separate the assessment of content and tools as being two separate evaluations. This minority perspective was especially interesting in light of the fact that the majority of participants had discussed evaluation criteria of tools and content as if they were separate aspects of a project. The minority argued that the evaluation of content and tools could not be divided because they closely influence each other and jointly develop the scholarship of a project. Participant T pointed out that though content and tools are usually evaluated separately, many "tools are built *for* a project. So you're kind of evaluating both [content and tools] at the same time." Another participant explained that the evaluation has to come by addressing the usefulness of the project in its entirety: "there's the whole design and architecture

of the site, there are visualizations and other tools and so on that have been incorporated into them…[and] *the edition is just a piece of the larger project*." [Participant P]. Separate criteria would diminish both the content and the tool evaluations by failing to consider the comprehensive influences from one aspect of the project on the others.

The minority opinion also suggested that digital scholarship was less a question of where to assign evaluation percentages (e.g., 40% tools, 60% content) than it was about demonstrating that the scholar was asking the right questions and could prove that that there was a scholarly methodology for *both* content and tools. Participant A expressed the following with regard to this symbiosis:

> [Evaluation] has to be both [for the content and tools]. You cannot do a good scholarly project without a content that is worth investigation…On the other hand, it cannot be a good scholarly project if you are using the wrong tools, if you don't develop them, the ideas, in the right way, if you don't ask the right questions. So methodology *and* content are what makes things scholarly (emphasis mine).

According to this statement, the tools, the interface, the content, etc. are all a part of a whole; each inextricably influences the other. Participant A's identification of the tools as part of the methodology of a project attempts to redefine their role in determining what is both useful and scholarly. Rigorous application and understanding of the ontological implication of tools allows the scholar to demonstrate their theory about the text and further refine the their critical approach to the materials. Without such methodology (and the integration of tools in this process), the content of a digital project may be seen as less successful in presenting a critical view of the materials. As such, it is possible to evaluate the different components of a digital project as part of the repertoire of scholarly activities. While the statement 'scholarly activities' is vague at best, there is a relevant consideration at its core, one that is fundamentally concerned with the definition of what constitutes scholarship. The community has to consider the impact of separating the evaluation of 'scholarly value' from the way that a digital

project is done, from how the tools are used, or how the data is generated. As expressed by the majority of the participants in response to the question, evaluation is a tug of war between the content and tools, with the different factors in opposition to one another rather than as facets of the same entity. Yet the comments by the minority about the linked nature of the content and tools underscore the point that scholarship cannot automatically be separated into discrete categories.

The responses to the first set of questions clearly emphasize that the creation and impact of a digital project do not exist in a vacuum. Digital scholarship has to be able to demonstrate the ability to influence people, practices, or future scholarship in order to be considered 'useful'. However, participant responses to what is evaluable about digital projects and the various definitions of what is 'useful' have raised a number of important points. The first is that there were few concrete factors expressed about *how* we would determine or measure the usefulness of a project. Obviously it would be impossible to determine what is useful without applying specific and articulated criteria to the project. Another confounding factor was that it was unclear which of these types of 'usefulness' was, as it were, the *most* useful and therefore should take precedence in an evaluation. For example, if the content of a project has an immediate impact on a scholarly field but does not allow for open access to it raw data, has that project failed in some respects? Also, 'failure' in a digital project is in some ways as difficult to determine as 'success.' As has been noted in previous scholarship, it is precisely the recognized failure of some projects that actually leads to more interesting discussions and innovative scholarship. Indeed, failure can even be more valuable if it pushes the boundaries on our current definitions of what constitutes scholarship [19]. Yet while failure is able to broaden our understanding of what is or is not possible for digital scholarship, how should it be evaluated? Traditional evaluation criteria rarely values failure, or at least, it does not encourage the repetition of the mistakes that are the grounds for making a failure interesting or informative to the community. Finally, as discussed above, the

'usefulness' or value of a project is often relative to a specific community. Participants argued that different aspects of a project made it 'useful', such as access to data, pedagogical usefulness, sustainability of the technology, etc. These factors were in turn connected to how a project demonstrated scholarship and therefore its value. But can 'usefulness' and 'scholarship' be used synonymously? And how do we attribute potential future value of a project to evaluations of its scholarly impact in the here and now? At the moment, all of these questions are open to active debate, the resolution of which greatly impacts the future evaluation criteria for digital scholarship.

B: Are digital tools scholarship? How do we measure this?

One point of interest from the first set of questions was that the majority of participants discussed the evaluation of tools and content separately. Indeed, more response time was spent overall in defining the 'usefulness' of one or the other without commenting on the relationship between them, nor on how the scholarship of the project was affected or enhanced by digital tools. However, a minority group of participants pointed out that it is problematic if evaluation criteria fail to consider tools part of a project's scholarship. As such, the next series of questions were designed to clarify how participants specifically evaluated tools and whether or not they regarded tools as scholarship. If participants did identify tools as scholarship, an additional question was posed on how the tools demonstrate this scholarship in terms of evaluation criteria. When speaking of tools, some participants had even used words and phrases such as 'flashy' and 'data collection' to describe the value of a tool. The principle research repeated these same terms when posing the next series of questions with the intent of inviting participants to reflect on their opinions and to consider the figurative association of their word choices. This strategy had the added benefit of avoiding the introduction of new terms that might bias participants' responses.

When explicitly asked about the place of tools in a digital project's scholarship, the majority of participants expressed awareness that creating tools requires a significant amount of theory and development which contributes to the value both of the individual tool and the overall project scholarship. Participants were divided on whether a tool was scholarship. Some responses associated tools with scholarship, but only in context to a project. Another set of participants argued that tools were definitely scholarship, both in relation to a project an also independently. Notably, participants who favorably argued that tools were scholarship were also self-defined tool builders. These individuals provided the most concrete evaluation criteria for tools and were also willing to measure it as such in a project evaluation. Yet even if participants did not feel that the tools were explicitly 'scholarship', all participants agreed that tools used in a digital project needed to undergo some form of assessment in the project's overall evaluation.

A number of participants did not consider tools to be scholarship in their own right. Instead, they viewed tools as facilitators of scholarly output. Descriptions of tools in this vein suggested a relationship to the content akin to artistic media: while the canvas and paint facilitate the creation of an artistic vision, the paint itself is not the 'art'. One participant made this comparison between tools and data collection in relation to the work for the Human Genome Project:

> I think that the Humanities are in a stage kind of akin to what happened with the Human Genome Project in that there was some science that went on in mapping the human genome, but most of it was just data collection. Very important data collection, but still, that's just what it was. People developed tools to collect that data, and those were important contributions. But all of that effort was essentially building up material that people could then use to do real science. You know, in finding cures for cancer, and so on…we're kind of in that stage ourselves in that we're putting a bunch of content out there that just isn't there yet, and …we have to do that in order to get to the next stage, which is really taking advantage of the semantic web and all the possibilities that it gives us [Participant I].

According to Participant I's perspective, it is not necessarily problematic to approach tools as facilitators of a larger concern: that of making material accessible

for other scholars and of generating raw data or material for use in later projects. Indeed, by his argument, this is still an expression of scholarship because it produces new knowledge. New knowledge allows for future discussion and engagement with the content. However, the implication is that the *data* is the important outcome, not the tools themselves, and therefore tools and their impact on the content should be considered secondary factors in an evaluation of a project's scholarship. The tool was thus seen by participants as intimately tied to the content and to the generation of new data.

Some participants considered new tools to be a legitimate form of scholarship. When questioned, Participant B said, "I think the tools are an output by themselves. And they should be the object of an evaluation themselves, *if they have been created by the project* (emphasis mine)." Another participant agreed with this perspective and added, "writing [your] own tools to allow for extraction and transformation of the text that you are working with…is a good and interesting area of scholarly activity in itself." [Participant D]. In both of these statements, the tools in a digital project are demonstrable scholarship. Notably, these participants referred to the creation of *new* tools, not necessarily the use of existing tools. When questioned on this point, participants assigned different values based on whether a tool was new or repurposed for a project. The value that participants attributed to new vs. old tools varied widely. If a participant was aware that a tool had been built for a specific project, they were more likely to see the building and use of that tool as scholarship.

The perspective that tools were scholarship was related to the emphasis placed on project methodology and the need to understand the ontological implications of the tool. Thus, participants did not see anything problematic in the use of an existing tool if it was able to contribute and enhance a project's goals. One participant explained that:

> There are different ways of doing things. You can build new [tools] or build new content using other tools, and I think both perspectives are very interesting. From a digital humanities perspective, I do believe that building

> new tools is what we are supposed to do. But…the fact that you're using technology that is on the cutting edge that was invented by someone else is not really an issue…the next stage in digital humanities is not building anything from scratch but using some of the things that have been developed in imaginative ways….inventing the wheel is not scholarly…[what is scholarly] is to move forward using things that have already been developed. [Participant A]

In Participant A's opinion, tools can be reused in innovative and interesting ways that may demonstrate scholarship, specifically by placing the tools within the methodology and critical apparatus of a project. There is even a possibility that a tool that is not created by the project may require more thought and planning to use in order to contribute to the editorial theory and methodology of the text. Participant D suggested that a project "include a methodological discussion of how [you] made those tools, why [you] needed them…and why there isn't an off the shelf solution to do that." Participant D pointed out the fact that simply building a new tool does not automatically confer value on that tool or project. Instead, a tool's value lies in challenging projects to consider what is already available for use and how to adapt that tool to a specific project. Reusing tools also diminishes the pressure to reinvent the wheel by creating new tools that address similar issues as existing tools. Evaluation of a tool therefore depends on primarily understanding what role the tool plays in the creation and output of the data. Participant F explained that, "If using the tool is part of producing what gets evaluated…I think what needs to be evaluated is whether you understand the tool." Scholarship is not only about identifying the most important questions, but is also concerned with addressing these questions in such a way that the knowledge and the data is applied in the right way, whatever this may be.

Based on these responses, it appears to be less important in evaluation if a tool was created from scratch or if it was adapted from a previous version of itself; what is important is whether the tool enhanced the project and is able to influence later tool design and implementation. As Participant T explained, "[if] you built this thing and someone besides yourself is using it…that is this huge validation of the

tool, that it wasn't just for that situation. It can be used by someone else, and it *has been used* by someone else (emphasis mine)." In the interview, the researcher and Participant T outlined four ways that reusing a tool confers value on both the project and the tool itself: First, there is the possibility that using the tool in ways other than it was designed may expose flaws that can be fixed and therefore would improve the tool. Secondly, the purpose and design of the tool may also be expanded to include these adaptations and be made available to the benefit of future projects. Additionally, the adaptation of an existing tool functions as an example to other projects of what is achievable using resources that are currently available. Tools also accrue value for money when they continue to improve and influence future digital scholarship. With this in mind, the use of tools is similar to demonstrating scholarship through a literature review. Just as a scholar shows due diligence in their scholarly research of the content and then building off of these materials to create their argument, building a tool from scratch may not demonstrate awareness of what has been done already.

Despite the extension of the usual due diligence expected for traditional projects in the creation and application of tools, participants expressed concern that traditional evaluation criteria are not equipped to accurately assess the value of tools. Participant D explained that while "I have a bias towards seeing the creation of tools as intellectual and sometimes scholarly endeavor", academic evaluators would not share the same opinion of tools: "a thesis committee or something like that would never think of [tools as scholarship]." The same concern was expressed by another participant, who recognized that "how you count [the value of a tool in evaluation] is problematic….[thus] there should be built in to the whole process an end point evaluation of the tool itself" [Participant I]. Participants suggested that tools are notoriously difficult for evaluators to assess; there is frequently a bias towards seeing tools as the novelty of a project but not its substance. This perspective is not limited to a faceless body of evaluators; some of the participants

had also used words like 'shiny' to describe the more technical aspects of a digital project, and implied that it could be a distraction from the content of the project:

> I think that the evaluation certainly needs to take into account [that] there has to be content under…the surface. You might have a digital project that's very flashy and aesthetically pleasing and you know has a lot of bells and whistles, but if there's not a whole lot of content underneath it that's original or…producing something of scholarly importance for the first time, then it's not really a good project. [Participant L]

Descriptions such as 'flashy' and 'bells and whistles' referred to the tools, interface, and usability of the project. There is a connotation of superficiality expressed by Participant L; that tools are not the 'meat' of a project, or at least cannot replace the content. This perspective prompted consideration of why some of the participants also viewed tools in this way. Two possibilities for this view were considered. First, it is true that many tools do not have the same expected lifespan as the original content. Tools are frequently subject to changes in technology, whether through upgrades or eventual obsolescence. The content, however, remains more stable; it may be derived from books or text files and is less subject to the same level of change that naturally affect tools. Therefore, tools are more ephemeral by nature, and may be less valued because of their transience, especially in a culture where print has long been the dominant media for storing information. In some respects tools may also be harder to assess because there is a less concrete analogy with the print world. Many of the participants explained that content was assessable because evaluators knew what to look for; the rules of evaluation were 'well-established' concerning critical editions and monographs. However, tools do not have one set of evaluation criteria that is well defined and consistently applied across the whole of the digital humanities, simply because tools come in a variety of shapes, sizes, and have different purposes and roles within projects, etc.

It became clear from the participant responses that the criteria for evaluation of digital tools differs widely across fields of study. Participants came from many different fields, including computer science, literature, linguistics, and library

science. Traditional humanities are not historically well versed in the evaluation of tools. Rather, such evaluation has long been the domain of fields like computer science, information science, or marketing. These fields, however, do not use assessment criteria that are accepted by the traditional humanities or used in the same way. Their understanding of what constitutes 'scholarship' also does not always encompass the same criteria that are valued by humanities scholars. For example, several participants who were not traditional academics advocated that the evaluation criteria should revolve around judging the effectiveness and success of what the tool was built to do. One of the participants who had a computer science background suggested the following criteria for evaluation:

> I look at how well the tool actually addresses the needs of the project. And a lot of time you can see great big mismatches. And sometimes that happens when tools were built around an original project and they're applied to something else…you can have very successful tools that are just sort of applied to the wrong sort of material to create a project that is not necessarily successful. The other thing I look for [is] actual users using the tool, the number of pages that have been produced using this [tool]…[and] whether or not it's been used outside of the project it was developed for….that means that there is a community around it, [it] is going to be sustainable, it's something you could recommend to other people…there's also the number of contributors. So how active is the community working on something? [Finally] there's also this question of 'what are your resources?' And every institution is different, every project is different, and so I don't think you can make a hard and fast evaluation criteria [Participant T].

In this scenario, a tool would be evaluated on the appropriateness of its application, extent of use, community support, and what was accomplished given a set of original resources. In the humanities, criteria tend more toward evaluating the appropriate application of a theory of the text under assessment. The application of such criteria to the evaluation of tools is much more frequent in relation to computer or information science projects.

Participant responses demonstrate that the idea of what constitutes scholarship needs to expand in order to consider digital tools and how to evaluate them. Because this requires the articulation and application of different evaluation

priorities, it was no surprise that many participants expressed discomfort with assessing the value of tools. In particular, participants were especially wary of evaluating tools that used components that were unfamiliar to them; the concern was that they were not 'expert' or qualified enough to perform an accurate assessment. Participants also explained that the ability to evaluate a tool required a high degree of knowledge about tools and tool building that would exclude a large portion of the community. While it is true that tools may appear inaccessible to scholars who do not use, build, or adapt tools themselves, participant responses suggest that it is also a challenge for scholars who *are* familiar with these activities to assess tools. Other participants noted that their own tools had been poorly reviewed for the same reason: that evaluators simply did not understand the goals and structure of the tools, and were therefore unable to fairly assess the tools and their role in the project.

Based on the answers received to this set of questions, a tool cannot *be* evaluated unless the criteria for evaluation are also clearly defined. As such, perhaps the evaluation of tools depends on the purpose of the evaluation. While it is true that digital tools are less readily understood by evaluators who do not know how they are built, let alone what they are meant to do, a fair evaluation of a tool must take into account these aspects and also demonstrate awareness of how the tool impacts the project. Thus, when one participant was asked if tools had to influence the scholarship of a digital project, and answered "not necessarily. It depends on your project topics [and] aims." [Participant B]. A solution that allowed participants to work around this dilemma of how to assess the scholarship of tools was in allowing the project creators to define what was scholarly about the project. Such a definition could include a case that the tools used in the project should be evaluated as scholarship. As we will see, several participants suggested that tools should be considered part of the scholarship *if outlined as such by the scholar or creator(s) of the project or tool who are submitting their material for evaluation.* Therefore, the tools should be able to be evaluated as an integral part of a project

(and perhaps separately) if the creator makes this request. Such a perspective is important and means that evaluation depends on first determining what the tool was meant to do, the purpose of its evaluation, the weight that the evaluation criteria carry, and finally who should perform the evaluation.

C: Who should evaluate digital projects?

The third series of questions focused on asking participants to identify what person or persons are the appropriate evaluators of digital projects. As the two previous threads indicate, evaluation of digital projects depends greatly on how the participant identifies evaluation criteria in relation to their definition of scholarship. Many of the participants had previously had digital projects evaluated or had been asked to assess other digital projects, and therefore had relevant feedback on the subject of who should evaluate digital projects. The questions were not only designed to identify the person or group that participants felt were appropriate reviewers for digital projects, but also to outline the necessary qualities associated with these persons or organizations. Finally, participants were also asked to clarify why specific qualities were desirable in an evaluator.

The person or body deemed most appropriate by participants for evaluation of digital projects was intrinsically related to what aspect of the project was being assessed and why. Thus, while all participants agreed that peers must review digital projects, the identification of specific groups and the desirable qualifications of these bodies varied widely. Some participants felt that evaluators should be discipline-specific peers or colleagues within institutional tenure and promotion committees. Others suggested that evaluators who were familiar with traditional print-based evaluation criteria were the best choice because print criteria have been a long-established mechanism for scholarly evaluation and provides an evaluative frame of reference. Another group of participants suggested that evaluators did not need to use traditional print-based evaluation criteria because there are significant differences between digital and print projects that require an evaluator to have

separate kinds of expertise. Participants also agreed that the goals of the evaluation directly affected the qualifications of the evaluator. If a project was up for a tenure or promotion evaluation, the type of reviewer identified for evaluation was an individual who could demonstrate expertise in the project's subject area and was employed in a traditional academic position. However, several participants agreed that someone with technological expertise should be included as a project evaluator regardless in order to ensure fair assessment of a project's digital components.

While discussions of evaluation criteria have made reference to the profile of the ideal evaluators that should assess digital projects, the problem lies in trying to determine which definition of 'scholarship' takes precedence in an evaluation. If 'scholarship' is seen from a field specific perspective, then the identification and evaluation of scholarship is left to the peers and experts of that project's field. One participant commented, "I think [the evaluator] is going to have to be discipline-specific, because different disciplines have different standards and value projects differently" [Participant I]. In this statement, Participant I has identified the problem of priorities. Individual fields have their own needs and perspectives on what constitutes scholarship. The assumption inherent in field-specific evaluations is that people within a given discipline inherently recognize the merit of a project in their specialty area *because of their training* in a specific field. As a matter of course, people outside of that discipline will not have the same training and therefore will not be able to evaluate the project in line with the specific priorities of that field. In some ways, this opinion is akin to a guild mentality, where one craftsman is able to inspect another's work, and someone outside that guild is not recognized as an appropriate evaluator. However, there are two additional implications made by this association. The first is that it assumes that people within a given field agree on evaluation criteria. The second and related assumption is that members of a discipline consistently apply the field-specific evaluation criteria. These implications are especially prevalent in academia, where training is designed to prepare scholars to add to work in a field. However, does this training also

adequately prepare scholars to assess the work of others? As Participant I explained later in the interview, "whether it's digital or analogue…the evaluation standards shouldn't really be different because if it's scholarship it's scholarship." Note that the definitions of scholarship are not clearly defined, only assumed to be inherently recognizable. Finally, while discipline-specific peer review has long been the route of traditional academia, this perspective tends to favor the content of the project over the digital components, and does not take into consideration the specific concerns raised by evaluating digital tools, as we will see.

Because there is no single set of guidelines currently used for digital projects, several participants suggested that an evaluator should be someone who is familiar with the traditional peer review and print publishing system. The reasoning provided in support of this perspective was that the standards for review are well established for print-based materials and transcend discipline-and-institutional specific boundaries. As such, the current evaluation standards for print materials provided a framework in which to apply evaluation criteria to digital projects. Participant H explained that, "We all know what a traditional printed edition looks like", with the implication that the community knows what an edition *should* look like. Another participant agreed, and explained "We've been taught scholarly concepts which we then apply to the medium that we're used to…and we apply our concepts to the evaluation of a print medium and then make scholarly judgments based on that" [Participant D]. Some participants did not differentiate between the kinds of questions asked by an evaluator for a digital vs. an analog project:

> the questions for print and digital projects are similar: how much is there that's original? How much new stuff is there? How easy is it to understand and find what you're looking for as a scholar or learner coming to this project? Same way if you're reading a book: how well are the arguments laid out? Do things go here? Does the index point you to the things you need to get to? *I don't see that those questions are really different in kind* [Participant L, emphasis mine].

If the questions between print and digital scholarship are indeed the same, then it would be appropriate to consider using current print-based evaluation criteria to

assess digital projects. The benefit of doing so is that evaluators would know what to expect and what questions to ask. There also would not have to be any additional or special training for the reviewers, because the underlying assumption is that a reviewer who is well trained to recognize scholarship in the print medium should be able to do so in another medium.

In the same way, participants felt that the evaluation of editorial practices and procedures would also be the same from print to digital. For example, Participant B said:

> I don't think the presentation of the work is quite different from the digital to the book. I don't know what your feeling is about that, but you still need to have an introduction about the manuscripts, your choices as an editor, the rest is just the digital world amplifying the possibilities of a book, but otherwise, I don't see how the standards would need to be so different, from the one to the other.

What Participant B's statement expresses is that she believes the methodology of creating a digital edition is the same used in a print edition. There is an association here between seeing the editorial practices as the scholarly nucleus of a project, while the digital components of a project are more analogous to the idea of a medium or a format. Participant D agreed with Participant B's perspective, and added the following on the evaluation of any edition regardless of its medium:

> The thing that we're not in any confusion about is the scholarly end of things, because *the same sort of scholarly rigor that applies to a print edition applies to a digital edition.* So you know, good editorial transcription techniques and principles of editing and citing of other witnesses…that kind of critical apparatus of stuff, providing details, introductions and investigations of the original witnesses, etc.—all of that … is required for a scholarly edition of any sort, *whether it's print or digital* (emphasis mine).

According to this perspective, the "scholarly rigor" of a print or digital editions can be assessed with the same evaluation criteria if the practice of editing does not fundamentally change between mediums [19]. This perspective also assumes that the reviewer will identify (correctly) the value of a digital edition by using the same set of editorial questions that would be relevant to a print project.

Despite participants' agreement that scholars who are familiar with traditional print criteria could apply these standards to digital projects, a number of concerns have been raised in the field about whether this is entirely effective or appropriate for the assessment of digital work [50]. By approaching digital editions as if they were simply extensions of print editions, the purpose and innovation of the digital edition may be overlooked [9]. In some respects, applying print criteria to a digital project is related to the kind of evaluation that is done. Participant B suggested that having digital projects reviewed by someone who is not familiar with digital tools may be useful in terms of a usability assessment:

> it might be good, because if… digital scholarship is accessible only to people who are digital savvy, maybe it doesn't make its job. If it's not usable or understandable in its entirety to people from the field, then, there's a problem. It should be, of course, it won't be evaluating the interest of the digital aspects of the project, but if it's not the point [of the evaluation], it should still be able to understand whether the scholarship is valid or not, from the field point of view [Participant B].

This perspective suggests that the point of the evaluation changes the profile of the evaluator. If the project is being evaluated for its "digital aspects", then the evaluators must be familiar with the technology to make an accurate assessment. Interestingly, Participant B suggests that the scholarship can be evaluated without an understanding of the digital components of the project, which hearkens back to the answers given in the first and second set of questions about the relationship between evaluation criteria and definitions of scholarship. However, trying to determine the kinds of evaluators required for specific kinds of evaluation is also problematic, and one that is an entire paper in itself. For now, the point remains that the aspect of the project that is identified as the 'scholarship' determines what kind of expertise the evaluators who assessed that project must have.

Over half of the participants recommended that the evaluative group or person must demonstrate an understanding of the digital components of the project regardless of whether or not it was also considered part of the scholarship. Participant D said, "I would think that it needs to be a set of people evaluating the

academic content of whatever's been produced…and the use or development of tools that have gone along with that." Another participant included familiarity with digital tools as part of the necessary qualifications of project evaluators:

> At a minimum, the roles you need…[are] someone who understands the subject matter…a subject matter expert…you need somebody who is…either interested in data curation or someone who has studied data curation to take a look at it from a 'can this stand' point of view…and someone who understands the encoding of the project…[Participant G].

Evaluation of the digital components of a project requires a "more specialized evaluation" than evaluation of the subject matter; at the very least, this evaluation is a different type of specialization [Participant A]. Just as traditional print-based criteria recommends that evaluators be experts in the content of the material that they review, participant responses suggest that reviewers demonstrate expertise in the technical areas that they are asked to assess.

The previous set of questions about digital tools had suggested that digital projects have to be evaluated in a way that includes both tools and content, especially if the tools are part of the methodology. By applying print-based evaluation standards to digital projects, there may be a higher degree of privilege afforded to the content at the expense of the digital tools. Trying to apply print-based evaluation criteria may also become problematic when evaluating the project as a whole rather than as separable components. If the digital features of the project can be said to influence the scholarship, then it is important to have both a content expert and a technical expert evaluate the project. It appears that to fairly assess the project it is necessary to include an expert in digital tools. Yet bringing in experts in digital tools poses another set of difficulties. Even though the situation is changing, several participants noted the difficulty in finding people who were experts in both tools *and* content. . As Participant A explained: "[evaluation] is very complicated to do for digital [projects] because you need to know the content and the tools and the data technical infrastructure, so it requires a 'super-user' to be able to do the peer review….". Based on this response, relying on a "super-user" is

problematic simply because these users are not always available. Participant T explained that, "something like the Digital Humanities is such a big field that no one can know all of it. And I could say, 'yes! We should singlehandedly evaluate all of it,' but I can't do it and I don't know anyone who can." Indeed, there were several participants who demonstrated a hesitancy to review the digital aspects of a project, despite the fact that they regularly use and create digital tools in their own work.

This tension between the need for evaluators to be experts in both content *and* tools is not a new one. In traditional departments this has been a longstanding problem, since there may not be other scholars within an institution (depending on its size) that understand or use digital tools. Many participants referred to NINES and 18thConnect as examples of oversight groups that attempt to evaluate digital projects in terms of both content and tools. Evaluations by these groups is done by "people who are experts in the content of the project….they'll bring in some people who have...the technical expertise to look at issues like usability of the site and information architecture and TEI encoding" [Participant P]. The review is then brought together from the individuals who assess the content and those who assess the technical aspects of the project. In some ways, such a model was represented as the most ideal type of review body for digital projects.

From these responses, it is obviously of key interest to identify what aspects of a project are evaluated and why, because of the direct effect on the qualifications of the evaluators. In the opinion of several participants, the argument for applying print-based criteria to digital works has multiple strengths: it is already well known and has been used by the scholarly community for years, it has clear and established criteria for evaluation, and finally, it is able to move more freely between disciplinary boundaries. However, is this in fact true? It is not clear whether print-based criteria should be the point of reference for evaluators of digital projects, despite the fact that these beliefs are widely held by the scholarly community. Are the criteria for print scholarship actually better defined and therefore evaluators who are familiar with this system are better qualified? Or is it simply the case that

scholars have received training in this evaluation system which makes them more comfortable and more consistent in the application of the standards that they unconsciously adhere to? Also, it is important to find out if digital tools can be accurately assessed using print-based evaluation criteria. Based on the preliminary data in response to this set of questions, it appears what is being evaluated and why changes the profile of who is best suited to evaluate digital scholarship. Participant answers indicated that an evaluator must be someone who can evaluate both tools and content assess a project, or if they are not qualified to do so, they must make this possibility available by consulting additional experts. To accomplish this, a group of experts, or a community of reference, would be valuable as a resource for both scholars and evaluation bodies.

D: What role do stated intentions play in evaluating a digital project?

After participants were asked to give specific evaluation criteria and to identify who should evaluate digital projects, it became important to determine how participants defined the role of the project in articulating its own evaluation criteria. Because all participants had either been involved in formal reviews of digital scholarship or had undergone review, the fourth set of questions asked participants to consider how reviewers and projects may benefit from a statement of intent. The importance of a statement of intent had been alluded to in some of the earlier participant responses. In many instances such statements were the only explicitly articulated criteria applied in a project evaluation. Participants were also asked to consider if a statement of intent helps to mitigate between a project's ideal evaluator and the reviewer at hand, who may or may not be the most appropriate judge of a project's achievements. While bodies like NINES have begun trying to meet the evaluative needs of scholars, many of these groups are still in their nascent stages and are therefore not yet fully functional. It was therefore of interest to see whether participants felt that a statement of intent was able to help a project make an argument for its scholarship, and what information it should include.

Ultimately, participants described the goal of the reviewer as facilitating an efficient and thorough review of the materials and providing constructive feedback about the project. To accomplish these goals, participants agreed that the project should provide assistance to the reviewer in the form of a document that outlined the intentions, limitations, and resources of the project. Participants indicated that this 'statement of intent' was the best way to fairly judge the materials because it used a project's own criteria. The statement was not only seen as providing the opportunity of expression of the editorial theory and methodology of the researcher, but simultaneously avoided the pitfall of trying to do a 'one size fits all' review. Participants also considered stated intentions to be the best way to evaluate the scholarship of digital tools, because the statement would allow the project to present a defense of the tool as scholarship by outlining *how* the tool contributes to the achievement of the project's goals. As such, the statement removes the burden from the reviewer to determine whether or not the tool is scholarly, at least in part. The statement of intent also allowed projects to be reviewed by a broader range of reviewers rather than a small group of ideal 'super users'.

It is a particular trait of Digital Humanities projects that they often feature a staggering amount of variation. Because there are so many different kinds of digital projects and editions, trying to develop a single set of evaluation standards is impractical. Also, using a single evaluation standard runs the risk of unfairly excluding projects that do not easily fit into definable categories. Happily, the fact that evaluation criteria need to be flexible avoids privileging orthodoxy rather than innovation, though it does not in the short term help evaluators to make accurate or fair judgments of digital scholarship. In order to address this problem, participants suggested that it would be more helpful to allow for multiple individualized evaluation standards within a larger articulated framework of guidelines. If the individual project makes its own goals and methodology clear, evaluators may then use this stated criteria for assessment. As one participant said:

> you want to have a situation where you can be evaluated on what you've done [and] also the standards that you've used…make the best case you can for whatever it is that you want to be evaluated on, and then I evaluate you on both the case you make and the parameters that you establish and being relevant [Participant Q].

Having a statement of intent partially relieves evaluators of the burden of trying to anticipate a project's main goals. The statement also gives projects a way to hold evaluators accountable for applying an explicit set of criteria. Another participant stated, "assessment should always take into account the…declared interests and purposes of the author. The creator of the markup or whatever…you should listen to what the person says that they were trying to do" [Participant O]. With a clearly defined and publicly available statement of intent, the evaluator is able to 'listen' to the project and avoid judging the project solely on the evaluator's own priorities. The statement also allows the project to make a case for scholarship to the evaluator(s). Participant F clearly expressed this point as follows:

> There are lots of different ways of producing critical editions, scholarly editions, and they have different goals and use different methods, and *in order to evaluate them you need to evaluate them for what they are.* It's not like there is a single standard for what constitutes a good edition except inasmuch…that the edition needs to declare its principles and its methods. If it doesn't do that, then you don't know how to evaluate it. If it does that, then you evaluate it against that declaration and see whether it appears to have used those methods consistently and looked up to those principles consistently. And I think that's probably a pretty good template for evaluation of digital projects, there are so many kinds, it would be meaningless to say 'all digital editions must [do X]. If they haven't [explained their methods] then you can say, 'well, that's not an intellectually responsible project, because they haven't done those things' (emphasis mine).

According to Participant F, it is in a project's best interests to give evaluators ample material to work with in an assessment. An explanation of the project's methods allows evaluators to check for consistency of application. A declaration of intent also provides the project with the opportunity to place itself within a larger framework of scholarship, to argue for its relevancy in the field and for its specific interpretation and treatment of the materials. If the project does not include a

statement, it is not demonstrating that it is taking itself seriously, or as F explained, it is not "intellectually responsible".

If a case can be made for digital tools to be evaluated as scholarship, then participants suggested that the evaluator should also have an explanation of the tool's methodology and role in the project's scholarship. Participant T explained that, in addition to best practices for documentation, explaining the methodology behind a tool is part of promoting future scholarship:

> Editorial methodology always has to be evident. I think that [projects] need to document how they're using the tool, their standards for data entry, and the tool itself needs to document when I see this kind of apparatus I'm generating these TEI tags… They need to expose, 'Here's how we're using the tool.'

T's quote is an interesting comment on the different levels of expectation applied to project evaluations. As expressed by participants in the second series of questions, a project must demonstrate an understanding of the tools used in the production of relevant data. This understanding is central to the overall methodology of the project. By including a statement of intent, the creator provides a teaching document that acts as a 'see one, do one, teach one' approach. In other words, the project demonstrates an clear understanding of how a tool works *by explaining it* and its output to others. As a result, the next project to use the tool will have a standard of reference for how the tool has been used in the past and how it was able to influence the project's scholarship. Thus, a statement of intent is as much a pedagogical tool as a demonstration of the project's scholarship because it provides reasons why a specific format or tool enhances or informs the project's scholarship. Such a demonstration can be used to represent best practices, both for future projects and for the application of a project's self-defined evaluation criteria.

E: Should the TEI be involved in the evaluation of digital projects?

Participants had thus far been asked to identify evaluation criteria for digital projects, the appropriate reviewer, and the importance of a statement of intent with

regards to a fair evaluation of scholarship. The next series of questions centered on an activity frequently involved in the creation of digital work: text encoding. As text encoding is often used in digital scholarly editions and projects, it is at a nexus point of both tools and content. The intent was to see where this specific activity fell in terms of the evaluation of digital projects. The purpose of bringing the Text Encoding Initiative (hereafter TEI) into the discussion was twofold. First, the TEI tag set is a long-standing and widely used standard for text encoding. The TEI has a large community that is active and interested in the application and development of these standards. Therefore, it was important to understand where text encoding falls in terms of evaluation and whether or not participants agreed that the TEI should be involved in evaluations. To clarify, the TEI encompasses many things: a set of users, the Consortium, the online journal, a set of tags, etc. [51]. By using the term 'TEI' in these questions, the intent was to refer to the TEI community, which includes all users. Specific reference to governing body (the Consortium, hereafter TEI-C) or to the set of tags created and maintained by the TEI is indicated where appropriate.

Participants were strongly divided on the role of the TEI in the evaluation process. The answers to these questions given were based on participants' understanding of the TEI's primary role and functions as a group. Many of the more experienced TEI users were not in favor of having the TEI community involved in the process of evaluating digital projects. However, other participants advocated that the TEI take part in evaluations of digital projects predominantly as a resource for others. Due to the complex nature of the answers received, the discussion will consider the reasons opposed, followed by the reasons in favor of the TEI's involvement in evaluation.

Those who were not in favor of the TEI having a role in the evaluation of digital projects believed that the TEI was simply not the most appropriate body to perform assessments. For many of these participants, assigning evaluative work to the TEI was a sticky wicket because it is primarily composed of volunteers. Others

felt that the TEI community would not be an appropriate evaluative body because it was not designed as a peer-review system. Instead, participants identified other bodies (especially NINES and 18thConnect) as the correct forum for evaluation of digital scholarship. These bodies are structured as a group of scholars who are experts in the content from a specific period of time and may also have some knowledge of digital formats and tools. Many also raised concern about the boundaries between the general practice of text encoding and the TEI specific tags. While TEI tags are widely accepted and used for text encoding, these are not the only method of text encoding or the most appropriate tags. Assigning the TEI community an evaluative role would potentially promote orthodoxy in the application of its tags and redirect finances from its other activities. Finally, participants identified one of the biggest problems with evaluation of text encoding, indeed, with evaluation of digital projects as a whole: namely, trying to define exactly *what* would be assessed in the evaluation and *why*.

One concern about the TEI community involved the political delicacy that would be necessary to develop an evaluative branch of the TEI, coupled with where the funding would come for such a venture. One participant explained the funding for the current TEI activities would make it difficult to implement evaluations:

> if they wanted to introduce a process of…assessment [that] will mean diversion of funding from other activities that they're doing...to do it right and to do it fairly. I don't think it can be done on the backs of volunteers within the TEI, or it won't be fair enough to actually do what people are going to expect it to do, which is stand in for tenure and promotion cases…and that means that they're probably going to have to do some form of cost-recovery…if they're not going to divert resources and if you want to have it done properly [Participant O].

Participant O suggested that there is a practical consideration for the TEI-C should it request or expect this kind of participation from its members. First, it is a non-profit organization, so it would essentially be asking people to give their time for free unless it could come up with a way to compensate evaluators. If it cannot, then it risks placing an undue burden on the volunteers. It may have to consider

withdrawing funds from its primary function, which is to "collectively develop and maintain a standard for the representation of texts in digital form" via "resources and training events…information on projects using the TEI, a bibliography of TEI-related publications, and software developed for or adapted to the TEI" [52]. The TEI-C would also have to consider whether it could justify the overhead that would be required to do evaluations, and whether there was enough interest for the TEI community to justify these costs. Thus, should the TEI get into the business of evaluation, it would require serious consideration about finding a revenue stream that would not tax the community's current time or resources.

Several participants couched their hesitancy about the TEI's role in evaluation in terms of the TEI's stated docket of activities and how the organization functions. The TEI's structure and hierarchy was not originally designed to act in a manner traditionally associated with print-based peer review bodies. For example, the TEI community is not comprised of people who are all part of the same scholarly discipline. It would be defined instead as a loose association of users bound by a common interest. One participant made the following comparison:

> what is the TEI, as it were? You know, it's a non-for-profit organization, there is a council, there is a technical board, but there are many, many people in the community who are interested in TEI, use TEI encoding, but are *they* the TEI, as it were? So it's kind of saying 'Well, the MLA should be in charge of something, [or] the MLA should participate,' but who is the MLA? A body of people who work in New York and a lot of members [Participant N].

Because the TEI community is in some respects like the MLA, it can suggest but cannot command how to uphold best practices of using TEI tagging structures. There is no real authoritative power; instead, its strength lies in offering suggestions to users based on the combined experience of its community members.

Other participants further explained that it was impossible for the TEI community to operate in the same capacity of a peer-review body like NINES, which meant that its evaluative focus would be of limited use. Participant P explained "There are different functions for different types of bodies, and I think that the TEI is

really good at maintaining the standard [of text encoding] and pushing that out to users and so on, and the reviewing aspects are better handled by groups like NINES." Another participant confirmed this perspective:

> TEI is not, or cannot, I should say, [provide] field-specific scholarly review. Because the TEI is fantastic…[but] if TEI were to give their stamp of approval to my project, and if I were to submit it to a committee for promotion, they wouldn't know what to make of it, because they're not in the field. I mean, it's in the field of digital studies…but evaluating scholarly relevance or pedagogical relevance is really a completely different story. So I would hope that these field-specific or era-specific literary humanities oriented review sites would actually solve [the reviewing] problem. [Participant S]

NINES and its associated bodies (18thConnect, and soon MESA) were seen as more appropriate for evaluating digital scholarship because these groups are more readily identified with traditional scholarship. According to its website, NINES identifies as a "scholarly organization" with three primary goals: 1) "to serve as a peer-reviewing body for digital work" in a specific time period (1770-1920) and in a specific area (British and American), 2) "to support scholars' priorities and best practices in the creation of digital research materials", 3) and to develop software tools to promote these activities [53]. The key point in this description is that NINES is a *scholarly organization*; that is, it is comprised of scholars working in a specific field and using digital tools. Participants believed that a NINES review was more likely to be accepted by academic committees as evidence of a project's scholarly value because the members of the editorial boards as well as the scholars solicited for reviews are recognized experts in the field that they are reviewing. In contrast, the TEI-C was not set up as a peer-review system, but as a community interested in a specific topic: text-encoding. Participants in traditional academia were less inclined to see a review of scholarship from the TEI as helpful because the TEI community is not an academic body or a demonstrable expert in a specific scholarly field. This does not mean that there is an absence of scholars who are also a part of the TEI; in fact, many scholars in NINES are also involved in the TEI community. However, the TEI-C as a whole was perceived differently than NINES

because its declared goals and activities centered on text encoding, not necessarily on a single area of scholarship.

For some participants this perspective on the TEI community and its capability to review digital scholarship highlighted the divide between traditional and non-traditional scholars. Participant P identified this problem in terms of the discrepancy between the role of the TEI and the role of peer review groups like NINES in offering an evaluation of a digital project's scholarship:

> there are many people who are excellent, really productive members of the TEI community and on the Board and Council who aren't traditional, who aren't professors. And I don't like to categorize people in those ways, but when it comes to something like tenure and promotion, you're going to have people who will put people into those sorts of categories. *I think you want the reviewing done by bodies that are made up predominantly of people in tenure or tenure-track positions to get over those biases.* If you could have three members from the TEI council who were doing a review of a project for someone's tenure, and they could be computer programmer, and a librarian, and a IT support person, all great scholars and TEI experts, but that *might not be convincing* to a more traditional person on a tenure and promotion committee, who wants those reviewers to be more traditional scholars. And that's the problem, I think, and I think that different people in digital humanities are working to legitimize the roles and qualifications of these alt-academic positions. But it's still a problem for people going up for tenure and promotion (emphasis mine).

There is a wealth of implications in this statement, but the most important is that there may be more inherent value attached to a review from a peer review group comprised of traditional academics than one that has been done by alternative academics or persons working in other fields (such as marketing or programming, etc.). Again, this returns to the fact that the TEI community was not designed to be a peer review body for an academic field. As such, these participants believed that it would be difficult for the TEI to perform evaluations without further work being done to "legitimize" or prove that the community members were able to demonstrate subject knowledge in addition to text encoding expertise.

The next difficulty that participants identified for the TEI is determining what parts of the project the TEI reviewers would evaluate. Because there are many

ways to edit texts, it is not always possible to determine what is 'better' about using a given set of tags in favor of another. Editing is both flexible and formative; two people may edit the same documents with different methodologies and produce work that demonstrates two equally interesting perspectives. The TEI standards of encoding were designed to accommodate the variations that occur in these editorial activities. Therefore, there is no absolute right way to encode a text beyond what will validate for a given set of parameters set up in the XSLT document that is outlined beforehand. Participant D questioned why there was even a genuine need to evaluate the text encoding for a project, simply because a given use of the TEI tags may not be generalizable:

> [one specific vision of the TEI is] not a generalized tool that somebody else could use unless they were doing exactly the same thing. So in that case, there's no real point in assessing unless you're wanting to assess me [and] how good I was at that job.

Based on Participant D's understanding of how text encoding would be evaluated, it appears that it would be the individual that was being evaluated instead of the project itself. Any attempt to evaluate individual preferences shy of an encoded document that does not render would be politically and practically difficult. Also, as the TEI tags are in some ways a very loose set of standards, it could cause problems of trying to enforce a kind of orthodoxy on how the TEI tag set could be used. In the words of one participant, this was in direct contrast with the original purpose of the TEI, which was "intended to be a community standard, not a 'from on high thou shalt' standard, rather a 'this is the way we all agreed to do this stuff' standard" [Participant G]. These responses make evident that it would be difficult to evaluate text encoding because it is unclear what aspects of the text encoding would be evaluated: the understanding of the individual editor of the TEI standards, their demonstrated use of those standards, the application of the TEI tags to further a specific editorial theory of the text, etc.

The problem of determining what to assess about the text encoding was frequently raised in tandem with a concern over the dangers of special privileging.

There were three kinds of special privileging that participants identified. The first problem identified with regards to special privileging has already been mentioned: orthodoxy in tag use. As we have seen, the TEI tags are deliberately very flexible to allow for a wide range of applications. However, practitioners who best know the TEI standards may also have preferred ways of using and interpreting the tags. These preferences are purely subjective and at present no one view of the TEI is necessarily endorsed in favor of another. However, some participants felt that, should the TEI community become involved in the evaluation process, one particular interpretation of the TEI tag set might be unfairly promoted:

> if you focus the work of evaluation in a single place, you run the risk of having people essentially beg the question of what's valid TEI by just…pleading essentially their own version by their own interpretation…[and] that would very much narrow the interpretation base [Participant Q].

If one person or a small group of people had the authority to decide what kind of TEI tags are the 'right' TEI, this could lead to other valid uses of the TEI standards being unfairly dismissed as problematic. As one participant explained, "Assessment is never neutral. It usually has some form of practice or method or technique, and it has stakeholders. It's done by somebody, to somebody, for somebody" [Participant O]. The TEI community would need to be very clear in determining *what* exactly was being evaluated about the encoding and what weight that carried for the assessment of the project as a whole in order to avoid unnecessary orthodoxy in tag use.

The second type of special privileging centered on the question of bias. If the evaluation of text encoding was undertaken by the TEI community, then it was possible, in the words of one participant, that the TEI would "have a sort of monopoly" on how text encoding would be evaluated [Participant S]. The TEI tags are not the only way to encode text, nor does the TEI community claim their tags are the best method in all cases [54]. If the TEI community were to evaluate text encoding, there could be a potential for hypocrisy since the TEI may be seen as having a clear stake in judging its own materials. Participant R noted that while

"there are people who are practitioners, familiar with various kinds of guidelines [they] also aren't *making* those guidelines at the same time" (emphasis mine). The difficulty lies in trying to determine how the TEI could judge itself fairly without potential bias.

Other participants noted that the third concern over special privileging is a danger of valuing text encoding over other facets of the project. Participant P cautioned that there may be more involved in a digital project than simply text encoding, and evaluation should take this into account:

> ideally the TEI is just one component of a successful digital project. That there are other things layered on top of that, you know, topic modeling, information visualizations, GIS stuff and so on, so having the TEI as the home for this is putting too much attention on one aspect of a project when you have all these other technologies involved that may be just as important as the TEI [in that project].

Digital projects may frequently use text encoding and the TEI tags, but these are certainly not the only 'digital' aspects of a project. As Participant Q explained:

> it's not 100% clear to me that you can't do good work that isn't TEI based. And so I would be a little bit nervous that you don't end up as an orthodoxy agency, and especially given that … the really really interesting stuff right now seems to be geospatial and semantic, and … also interoperability, so stuff with JSON and things like that, and I think it might be wrong to just choose markup [to evaluate].

The point here is a relevant one: why pick on text encoding rather than the other tools used in a project, especially if the text encoding is not the most interesting or innovative aspect of the project? Use of the TEI tags is not inherently indicative of a rigorous scholarly methodology, and text encoding is only one such way of formatting and exporting data. As such, participants argued that the quality of the markup could not act as a surrogate for a project's overall quality.

Despite these concerns, some participants maintained that there were several important reasons to have the TEI community involved in evaluation. Indeed, even participants who had originally expressed disfavor with the possibility of the TEI's involvement also noted that the idea still had enough merit to warrant

further consideration. The main points given in favor of the evaluation by the TEI were as follows: 1) that the TEI tags could be considered part of the editorial methodology that contributes to the scholarship of a project, 2) that the TEI community has a declared responsibility as a 'resource to scholars' to provide assistance if requested in this area, and 3) that by making the effort to review such projects, the TEI-C would provide an environment for further dialogue on both the use and function of the TEI tags as well as the development of evaluation criteria.

Participants who were in favor of having the TEI community evaluate text encoding in projects connected the use of tags with a project's editorial methodology. The decisions about how to encode the text in a given way were seen as affecting the scholarly quality of the project. Because of this connection to the overall vision of the project, participants reasoned that the text encoding should be considered part of the project's scholarship and evaluated as such:

> people, have argued that the act of encoding itself is an act of scholarship. Right? That you wind up dealing closely with the text and making editorial decisions that not only in and of itself is sort of good for your health, but that also is adding to the world of scholarship, producing new knowledge in a sense. Because you are providing your own analysis, and your analysis, while maybe not described in prose the way it would be in a monograph or a journal becomes implicit in what you do. Or maybe even make it explicit, but that it itself is a sort of work….[and] this is what you would want to be evaluating [Participant M].

Participant M has noted that there is a correspondence between editing, markup, and scholarship, or a 'theory of digital text' [55]. Participant S attempted to explicitly define the role of text encoding and tagging in the editorial methodology of a project:

> where do I think that tagging and text, the marking of a text, fit between editorial practice and writing an essay, or a book…[it] comes down to the level of granularity, I think, that you tag. Because in tagging, that's one of the main questions…. [so] the decisions that you have to do as to how much granularity, to deal with the text, is certainly, at least as scholarly as trying to figure out what kind of edition to write. And so it's in that process of deciding the edition, in the process of justifying an edition, in the process of

> figuring out the details that will go into the textual apparatus in an edition, which are, those are the most scholarly parts of an actual edition text.

According to this response, evaluating the text encoding of a project becomes a matter of granularity in an editor's approach to the material. An editor may decide to make these tags part and parcel of their theory of the text. To do so, the editor marks up the text in a detailed, methodical way, and the text encoding becomes an explicit part of their theory of the text. The problem therefore is not whether text encoding *can be scholarship*, but rather how it makes its case to reviewers.

Some participants felt that the TEI tags in the project should be considered in any fair evaluation of the work as a whole. Participant L compared the text encoding of a digital project with the editorial methods of a print edition:

> I think if you are evaluating the scholarly value of the digital project whose underlying data is in TEI, you really do need to look at that TEI. In the same way as if you were evaluating a letterpress edition…you would certainly be looking very carefully at the editorial methods section. You'd be looking at the transcription principles…to see that the conventions were applied consistently…if you weren't using that as a reviewer…you'd be leaving out half of what makes it a scholarly edition. And I think that's true of anybody who is claiming that they are making use of TEI to code for archival purposes. I mean, if it's very simple content and you're really just coding paragraphs and lines that might not apply. But if you're doing anything more sophisticated to capture things like textual variants and manuscript witnesses or semantic…or linguistic features of the text, then you…should be getting credit for the work you've done, and also if it hasn't been done well, then that should be part of the evaluation [Participant L].

In some ways, this response highlights the problematic position of editorial theory and practice and the value afforded these activities in academia. If participants saw this as a valuable outlet for scholarship and also identified text encoding as part of editorial methodology, then they were more likely to say that the text encoding should also be evaluated and that the TEI community was the appropriate group to do this.

Several of the participants advocated for the TEI community to be involved in evaluation because they felt that the TEI could help them improve their projects.

This perspective was predominantly expressed by participants who were less experienced with the TEI tags, whose work fell outside large and well-supported areas of scholarship (such as American or English literature), who were junior tenure-track scholars, or individuals in alternative academic positions. These participants were much more in favor of having the TEI-C provide either an evaluative review or a formative consultation for their projects. One reason for this support was that although the TEI-C already does a significant amount of outreach, providing formal reviews was seen as a resource to assist those who have beginning to intermediate knowledge of the TEI tags improve more rapidly. Several of the participants who are junior scholars said that the complicated nature of the TEI tags and the length of time required to become an expert in using it made it prohibitive for them to use beyond a beginner level:

> I feel that for someone like me who can't practice TEI all the time because I have other scholarly demands, I think that that becomes hard…. I think that in that sense I would want someone who was looking at the TEI to understand what the intentions were in order to kind of judge the scholarly aspect of it [Participant R].

Because their knowledge of TEI standards was minimal, junior scholars like Participant R were not able to fully exploit the TEI tag set to reflect their editorial methodology. As one participant said, the "primary place [of the TEI] is to serve the needs of scholars who are producing digital editions … of important source documents" [Participant M]. According to this perspective, the TEI community exists *as a scholarly resource* for a specific set of activities revolving around annotation and editing of documents. As Participant T explained, the "TEI exposes the process of editing, publicizes the process of editing." This definition is crucial, because it suggests that the value of the TEI standard is tied to its use, and that it may be a part of the larger methodology of editorial practice.

Due to the fact that there is still a need to define what constitutes digital scholarship to peers outside of the field, the TEI community was also described as potential sounding board for people who want to make a case that text encoding

demonstrates scholarship. One participant explained that there needs to be "a lot more work on the part of the scholar who's going up for tenure to educate people about digital humanities scholarship" [Participant P]. Many study participants felt that the generation of scholars trained in traditional (ie, non-digital) humanities did not recognize these activities as scholarship, and thus junior or non-tenure track faculty have a difficult time convincing colleagues of the merit of their work. Scholars in this position felt that the TEI could help them change the perspective through expert knowledge and also through dialogue with the community. As Participant R explained, "to make digital scholarship evaluable, part of what you need to do is talk about it a lot, explain it a lot, and in a sense, publicize it." Participant H agreed with this point and also expressed that the TEI's help would allow him to demonstrate the value of a current project to other evaluative bodies: "If [the funding bodies] knew the TEI and the quality of the encoding that is possible in this way…in order to have an idea of the real amount of work [they would] better understand our achievements." This additional work to show the merit of digital work as scholarship is where the TEI may be able to help demonstrate how a theory of the digital text might be represented practically and methodologically.

In an ideal evaluation, projects reviews would use criteria that fairly assess the project's work and that allow the project to demonstrate its scholarship to non-digital peers. As evidenced by participant responses to the fourth series of questions on the role of stated intentions in evaluations, such a review is possible only when generalizable criteria are placed alongside the stated aims and intentions of a project. Thus, if a scholar feels that their work in text encoding demonstrates scholarship and states this belief in their editorial methodology, the evaluators are duty bound to consider that argument. Participant O succinctly expressed this point: "wherever you stand on whether or not there's scholarship [in the project], it is still the case that we have people who are arguing that it is, and that therefore the TEI should provide guidelines for this." By evaluating projects, the TEI community would provide an additional measure of evaluation for projects that requested it.

Indeed, the TEI website (and as we will see, its online journal) provides an ideal platform in which to help scholars outline a philosophy of encoding or statement of intent, which participants had previously identified as critical in the fair evaluation of digital scholarship.

Other advocates of the TEI community's involvement in evaluation pointed out that while there should not be division between the different aspects of review, it is problematic to ignore the fact that there is no single individual who can evaluate every aspect of a project. Working in tandem with groups like NINES would allow for reviews to be both content and technically based, providing a resource for where to find qualified reviewers for text-encoded projects. Additional collaboration would also perhaps encourage dialogue between disciplinary fields if it was possible to see how others have addressed the problems faced in their own work. Participant R shared the following:

> I think that the interesting thing about when you learn TEI with different scholars from different periods and different disciplines is that you start to learn what's important to them, but that doesn't mean that it doesn't leak into your scholarship… sometimes it would be helpful to know what another discipline or another period is going to call attention to, because then you realize, 'Oh, well my author was actually…reading off of all kinds of dream vision texts, so …if I could understand what someone was paying attention to who was marking up medieval dream vision texts, then that actually might help me … think about that. So in another words, it's a different means of collaboration across boundaries, which is ultimately one of the things that digital humanities does really well. And so I think that it would be really great to have a body of people who are vetting all kinds of different texts, and then could say, 'Well, you made this choice, that's really interesting, someone who is working on something in a different period was making these kinds of choices, you might at least want to think about it.'

Participant R suggests that by providing access to review materials, the TEI would potentially be fostering additional collaborative work. Editorial theory and practice has always benefitted from exposure to the community; text encoding may also benefit, not only in advancements to the current tag set, but also in the scholarly understanding of how to interpret and theorize different kinds of texts.

Despite the fact that participant perspectives varied widely on the TEI's involvement in evaluation, there are a number of important points. The first is that evaluation of the text encoding cannot be a surrogate for the whole project in an evaluation. The TEI community is not equipped to evaluate digital projects in the same way as NINES because its infrastructure was not intended to function as a field-specific peer review group. Another consideration is that the TEI itself has a vested interest in encouraging its own tags, which may lead to orthodoxy or a bias in evaluations. However, there are also a number of compelling reasons why the TEI may still have a role to play in the evaluation of digital projects. Using the TEI standards is complicated enough that it would be a considerable asset for the scholarly community if the TEI community were to be an evaluative resource for people who need assistance. Though there are many stumbling blocks to determining the precise role of the TEI community in project evaluations, participant responses suggest that there are a number of relevant directions to consider with regards to the TEI and assessment of text encoding.

F: In what ways might the TEI community participate in the evaluation of projects?

The final question series was designed to follow up on how the TEI community might participate in the evaluation of projects. The main point of these questions was to discover if it would be practical for the TEI to evaluate projects and collaborate with peer review groups in a review context. It was also a priority to see what environments the participants thought would be optimal for evaluating the text encoding of digital editions. Finally, participants were asked to consider what documentation would be necessary for the TEI community to provide with regards to project reviews.

When asked what functional way the TEI could contribute to evaluations, there was a general consensus that the TEI-C needs to provide at least two reference documents as an evaluation resource. These references would be twofold; the first would be guidelines for review of projects. The second would be a list of certified

people who are capable of providing evaluation of digital projects that use the TEI tags. The website was seen as the appropriate place to post these resources, which would take two forms. In the first manifestation, references would be outlined as general guidelines that could be used in different evaluations. Participant I suggested, "we should be proactive in suggesting guidelines for evaluation because so many tenure committees and evaluation committees don't know where to start. " The TEI community includes all levels of TEI users, some who are experts in text encoding, while many evaluative bodies do not have knowledge of text encoding or a sense of how to evaluate it. As such, the documents would assist reviewers by promoting a better understanding of text encoding among the more traditional and less technically savvy review bodies.

The second resource that participants identified that the TEI-C could provide on the website was a list of certified individuals who would be capable of reviewing the TEI tag set in a project and advising on best practices. Participant L confirmed that knowing who is capable and available to provide assistance was crucial for first-time projects, and that it was the responsibility of the TEI community to "think about…things like certification, or coming up with a webpage that would list people who are considered qualified to evaluate" TEI encoded materials. Participant S explained that there is a "need for actual expert reviewers out there, or at least some sort of firewall that you can submit something to and have it get a stamp of approval, that not only does meet best practices for a digital edition which is a technical evaluation, but it also has some scholarly merit, or at least utility for the field" [Participant S]. Another participant expressed a similar opinion, going so far as to suggest that there should be documentation provided to assist in the creation of evaluation guidelines that would facilitate such reviews:

> I think the TEI would be better off first of all maybe commissioning guidelines for assessment. So saying, 'this is how we would assess it if you were a chair, or someone on a tenure committee, or whatever it is, this is…what we recommend you do as part of an assessment.' Sort of describe commission and the creation of a best-practices document. And then share

> that best practices document with the community. And then, possibly, if you were going to get into a formal certification… [Participant O].

Participant P also shared a similar perspective and added the suggestion that the TEI community might consider providing these materials to aid in the review work done by NINES and similar groups:

> they might offer as a service to other groups like NINES…[the] names of people from the TEI community who are willing to do technical reviews of projects. And NINES would have that resource to use if they…need someone to evaluate the technical aspects of the TEI encoding. [They] might go to that resource to ask for a volunteer to do that.

In Participant P's opinion, NINES and other bodies may benefit from having a list of approved reviewers for the use of TEI tags in multiple ways. First, it includes input from members of the community that are outside of traditional academic institutions. Such a list may allow scholars who would not normally have the opportunity to review such projects to be considered. As one participant explained, "If…you have a project and you want to make sure that your project is set up, the TEI is a … a mechanism by which experts are put in touch with projects that need advice" [Participant Q]. Having such a list of people would likely involve some form of certification process so that the evaluative body could verify that the people on the list were well qualified to provide suggestions and support to others. If the TEI-C provided a list of the names of people who could assist with these early and intermediate assessments on the website, it would also act as a resource to those members of the community who do not have a readily available or appropriate peer review body for their materials. The TEI community would not be taking the place of the peer review body, but it might be able to answer questions with respect to the editorial decisions and use of the TEI tags in that project. While the TEI listserv does already answer questions from the community, it is only able to answer questions that are posed to the list; there is no one who is necessarily looking at the entire project to see if the question posed is part of a larger issue with the work. Also, it is possible that numerous and different answers are given, and without

experience the petitioner may not know which answer is most appropriate for their situation and skill level. Therefore, it may be more beneficial for some projects to have someone look at the entire structure and provide feedback at length.

Participants also identified formative assessment and consulting on a project as a benefit for both the TEI community and the project under review. Reviewing digital projects early in the development and planning process as well as *medias in res* is potentially more cost effective. Also, receiving an outside perspective may help projects to avoid common mistakes in their infrastructure and editorial decisions. Participant O suggested:

> it should become a best practice to weave into [the project budget] some funding to bring someone up to consult, to get some sort of outside view, and so that's a form of assessment that the TEI could help with….the TEI might want to provide recommendations or ideas about this type of formative assessment for projects. Because these projects—earlier in the project, they soak up money. A good outside consultant can save you money…

Not only would a consultation potentially save projects money, but it would also allow the TEI-C to help promote best practices and innovation in the use of text encoding and the intersection with other tools. One of the criteria given for evaluation of digital projects was whether the project contributed additional knowledge to improve the tools that it used. As such, it is possible that these activities may also provide the TEI community with future improvements as more people continue to use, learn, teach, and refine the TEI through the evaluation of digital projects. It may also help project leaders consider their methodology and begin documenting their editorial decisions earlier. As demonstrated in the previous discussion, such assistance was exactly what several of the junior scholars and participants outside of large fields of scholarship had stated as a primary reason why they wanted the TEI to be involved in the evaluation of digital projects.

There was discussion from some of the participants that the TEI could also provide materials that would help individuals seeking evaluation to find the right arena of assessment. Participant P recommended that in addition to a the web page

about who can provide evaluation assistance, the TEI-C should also provide a 'how to' page of information that would feature common questions about how to get a TEI-tagged project evaluated and what would be assessed. The page would act as "a resource for scholars who need their work reviewed, pointing them in the right direction and [featuring] examples of projects that have been successfully reviewed." The TEI is certainly not the only group to perform evaluations of digital projects, nor is it always the best one. However, the TEI can still provide FAQs about reviewing text encoding (or digital tools in general) which may foster additional dialogue within the community and among the peer review groups.

Multiple participants identified the Journal of the Text Encoding Initiative (hereafter TEI-J) as the appropriate place to promote these review standards and suggestions for future work. The TEI-J is online and there is no access restriction; anyone can read the articles available. At present there is no dedicated reviews section to the journal. If the TEI-J was to include a section that provided a forum for discussing projects, it could publish and commission reviews of completed and ongoing projects. As one participant noted, the TEI-J could: "Make the review process itself useful to people, and that's where I would recommend the TEI Journal. It should get into reviewing. If they had a reviews column, then people who did [the reviews] could rewrite their reviews and submit them and get it published, just like a book review" [Participant O]. The TEI-J could decide whether or not to provide short reviews and links, or "more critical pieces that also deal with some other factors…like new uses of P5…[to give] people the space to be more thoughtful" [Participant N]. By providing reviews of digital projects, the TEI-J would function as a kind of "vetting" space, a place that will not only "get people up to speed from the initial understanding of the TEI" but also show the more experienced members of the community how to expand the possibilities of the use TEI tags in scholarly work [Participant R].

PART IV: CONCLUSIONS

This study addressed a number of questions relating to the evaluation of digital projects. Perhaps most importantly, the results confirm that determining criteria for evaluation is complex and nuanced. Due to the range and scope of digital projects, there does not appear to be a single hard and fast rule for evaluation that can be applied in all cases. However, the study results not only raised a number of interesting points about the evaluation criteria for digital projects, but also about the place of the TEI in the discussion of this topic.

Firstly, participant responses indicate that some criteria may involve areas of scholarship that are not always included in humanities evaluations. The idea of a project's 'usefulness' encompassed a variety of meanings, both in terms of the content and the digital tools used to display, explore, and analyze this content. Many participants indicated that 'usefulness' encompassed the ideas of usability, accessibility, long-term preservation, interface design, and the generation of open data. These concepts of usefulness have historically been the domain of fields such as computer science, information science, and user-experience design, not necessarily of the humanities. As such, the humanities may need to consider enlarging its criteria both practically and theoretically in the context of a larger issue: how we define scholarship.

The definition of scholarship relates to how a given field defines and articulates what is valuable to that community. Thus, while all participants agreed that a 'useful' project should contribute new knowledge, the definition and assessment of these contributions differed widely. Evaluation of digital tools especially highlights this spectrum. Participant responses suggested that tools could exhibit varying degrees of scholarship. However, while tools are products of scholarly activity, their value as scholarship changes based on a given field's point of reference. For example, several participants working in traditional English departments saw digital tools as secondary facilitators to the 'real' scholarship,

rather than as scholarship in their own right. In contrast, participants with computer science or tool-building backgrounds evaluated the same tool more highly as scholarship because of its innovation and user-friendly design.

Consistent with these divergent views, many participants divided the content from the tools in the application of their evaluation criteria. In this scenario, the content usually received the main attention of the evaluation, with the tools predominantly reduced to a secondary or facilitator role. For those advocating this view, the tool is valuable because it is usable, but this carries less weight compared to the innovation and intellectual rigor of the content. However, in many digital projects, the tools and content are not easily separable in the execution of a project's goals. If the content is poorly researched or uninteresting, then it does not provide the foundation for the tools to demonstrate their innovation or methodological importance. Conversely, if the tools are difficult to use, then the content cannot accurately demonstrate the tool's contribution to the textual theory or relevance in the interpretation of the produced data. As such, evaluation criteria need to take both content and tools into account, with particular emphasis on their relationships to each other and the goals of the project. In this way, the overall execution of the project is the final word in its evaluable scholarship.

Due to these considerations, this study suggests that evaluation criteria should be directly impacted by the stated intentions of a project. Explicitly outlining the project intentions and goals relieves some of the evaluator's burden of trying to determine what aspect of the project is the source of its scholarship. Regardless of whether or not the evaluator feels that some aspect of the project is scholarship, defining the project goals facilitates performance of a fair assessment in which the evaluators must seriously consider the merits of the project in accordance with the stated intent. For example, if the project statement articulates that a tool influenced the methodology of the project, and if they can demonstrate how it does so, then participants agreed that the tool played a larger role in the project evaluation. While active debate continues in this area, evaluation criteria must be flexible

enough to make allowances for stated intentions, which in turn enhance the larger discussion about what is scholarship.

This study also considered the role of the TEI community (Consortium and users) in the evaluation of digital projects. Because evaluation criteria depend on both stated intentions and the field-specific definitions of scholarship, there may be a role for the TEI to play as a community with expert knowledge of text encoding using the TEI tag set. As such, this study identified two activities for the TEI in relation to evaluation. First, the TEI should continue to provide counsel and formative assistance with ongoing projects, but in a manner that is targeted towards project evaluations. This service would ideally be available both to members of the general community and to other evaluative bodies, such as NINES. Secondly, by making provision for the submission and review of text-encoded projects, the TEI enhances the knowledge of the entire community because this process would likely lead to modification and improvement of its tag set. However, though the TEI may be involved in the evaluation of digital projects, the study results also indicate that there are definable limits to its involvement. It would not attempt to provide evaluation services in the same scope of NINES and other peer-review bodies. The TEI should also not attempt to enforce a single canonical use of the TEI. Instead, the goal of the TEI's involvement would be to provide an evaluative resource that helps scholars from a wide range of fields tailor the TEI to individual applications and to develop a theory of encoding that integrates with the overall methodology of each project. Practical recommendations for implementing these functions will be discussed further in Part V.

The salient conclusions of this study are that evaluation criteria for digital scholarship are anything but established. Experts in the field do not necessarily agree about what factors are most important in an evaluation of digital scholarship. Much future work will need be done in this area to unravel the current perspectives on what digital scholarship is and how it should be valued and assessed by the community. This study is therefore only the first in a series designed to further

clarify what we know and what we need to resolve in order to develop evaluation criteria for digital projects.

PART V: FUTURE DIRECTIONS

While there are still many unresolved questions regarding the evaluation of digital projects that invite further discussion, the results of this study have yielded a number of fruitful next steps for the future. The first suggestion is that the TEI-J might consider including a dedicated reviews section to feature project evaluations and accept submissions for review. Another apropos step would be the drafting of a whitepaper for the guidelines of the review of TEI projects submitted to the journal. This document would ideally be tested on several sample projects in order to demonstrate how these guidelines would operate in practice. After developing and refining a whitepaper, it would also be appropriate to consider how the TEI community might provide a list of approved evaluators and what kind of selection criteria this should involve. Finally, these findings also demonstrate that the digital humanities community has further work ahead in this area to clarify and define what criteria are valuable in the evaluation of digital project scholarship.

To facilitate further discussion of what evaluative reviews for text encoding projects might look like, we suggest that additional consideration be given to outlining guidelines for the submission and review of projects on the TEI website. Though the main TEI website has a project page that features a list of projects that use the TEI encoding schema, there are no evaluative features applied to the submitted projects. Instead, the submissions are mainly about visibility and promoting awareness of what projects have used TEI tags in their work. These submissions include a brief description of the project, the associated URL and host information, along with a contact address. While it is helpful to know what projects have used the TEI tags, we suggest that a new page be created by the TEI to expressly feature materials related to evaluation. Ideally these materials would be accessible both from the main and journal pages and would contain information relating to project reviews, formative assistance, the whitepaper, and examples of projects and statements of intent.

The website would need to include a whitepaper that outlines guidelines for submission and review. In order to draft this document, it would be necessary to speak further with members of the TEI community with the goal of determining the following: Who would want something reviewed by the TEI? What would the purpose or goals of the evaluation be? What aspect of the project is/is not being reviewed? How would the evaluator review the submitted project? Based on answers to these questions, it would then be possible to decide how to frame the guidelines in such a way that they were beneficial to both the reviewer and those submitting materials for review. These guidelines would then be compiled into a short document to be made available to the general community.

In addition to these guidelines, it would be apropos to demonstrate what materials submitted for review to the TEI-J might look like. NINES and its sister groups have a mechanism for integrating submissions into their interface and endorsing them as demonstrable example of scholarly, peer reviewed projects. Though it would not be necessary to publish the submitted materials or host them on the TEI's interface, maintaining a set of sample projects would provide concrete examples of what types of projects may be reviewed by the TEI community. It would also demonstrate the kind of evaluation projects should expect to receive if they submit materials to the TEI. These sample projects could also take the form of case studies in order to demonstrate the outcome of different evaluative goals.

One particular aspect of sample projects that should be included with reviews is a statement of intent. Example statements would assist in the fair and thorough evaluation of the project by explicitly identifying the scholarship of the project. The statements may even be of interest to the wider community in terms of what criteria are used and the weight that they carry. Therefore, we suggest that several examples of these statements of intent be made available for reference. A list of leading questions may also help projects to formulate their statements, such as: How did you outline your encoding schema? What choices did you make in the structure and analysis of the material? How did you apply the tags? Are there any

aspects of the content that you believe were limiting or challenging factors to your text-encoding scheme? What is your theory of the text that supports your use and adaptations of the tags? By thinking along these lines, the statements for submitted projects may help others articulate their methods and assist in the formation of what should be expected from an encoding statement of intent.

Two additional components could also be included as part of the evaluation resources on the TEI website: FAQs and checklists for reviewers and project submissions. Interviewees had suggested that a set of frequently asked questions would help the TEI community to define its own evaluative role and direct people to the most appropriate reviewing body based on the goal of the project evaluation. From the topics discussed in the interviews, we derived a number of potential FAQs that might include the following: What evaluative body should review my project? Is the TEI the right reviewer for me? When should I expect feedback from the TEI? What parts of a project does the TEI review? Why does the TEI evaluate projects? What documents would I need in order to submit my project for review? What will I receive in return from the review? How many reviewers will there be? Do I need to pay for the review? Do you have a list of people who can consult on my project? Will the results of the review provided to outside bodies? Who do I contact to ask for a review? Can I choose my reviewer?, etc. These questions may help to clarify what exactly the TEI's evaluative role would be as defined by the TEI community. Addressing these questions would also ensure that the TEI receives submissions that are in line with the kind of review that the TEI is willing to perform.

It may also be helpful to include a set of checklists as part of the review page resources to assist both the evaluator and the applicant in defining their roles and expectations. Based on suggestions received from the study participants, these guidelines would also be framed as a list of possible questions that the reviewer and applicant may apply in an evaluation, as demonstrated in Geoffrey Rockwell's model [56]. For example, a checklist of questions for the reviewer performing an assessment may include the following: Is it well formed? Is the use of tags

consistent? How does the project challenge or broaden the use of the current set of tags? Do they customize? Has a statement on the philosophy of the encoding been included? Does it meet the standards outlined for the project in the statement? Why/why not? Likewise, persons submitting for review may want to think about the ways they have documented their materials, whether or not their theory of the text is evident in their statement of intent and use of tags, and so forth. While these questions reflect only a sampling of what may be included as guidelines for the reviewer and submitter, they demonstrate the need for further consideration of and discussion about the evaluative criteria for digital projects.

Ideally, the reviews of TEI projects would take place within the framework of the TEI-J. In order to develop this possibility, the findings from this study provide two recommendations. The first suggestion is for the TEI-J to consider adding a section in its structure that would be dedicated to project reviews done by the TEI reviewers. This section could include both short reviews (between 500-1500 words) designed to provide a brief analysis and highlights from the project, as well as longer reviews (1500-3000 words) in the event that the reviewer desires to consider the project in greater depth. Members of the TEI community who are approved to evaluate projects could either volunteer or be invited to review submitted projects and report their findings to a reviews editor. The reviews of projects, including the special circumstances or ways in which the TEI was used, would then be accepted and published on the same timeframe as the rest of the TEI-J, unless it was necessary to increase/decrease the publication of reviews based on the number of project submissions.

To create the documentation for reviews in the TEI-J, including the whitepaper, it is imperative to engage in further discussion with the TEI community. Not only would the community need to consider who should evaluate projects (TEI volunteers vs. solicited reviewers, the TEI editorial board, etc.), but it would also have to determine what (if any) infrastructure changes may be necessary to pursue these goals. Concomitant with these concerns is how to compile a list of reviewers

who would be able to perform the evaluations. First, it would be necessary to determine who would receive and the review the submissions, and how many reviewers would be necessary for each project. Also, it is important to ensure that the review work is consistent, that overt bias is avoided and that no single perspective of the TEI tag set was given unfair preference. Therefore, the TEI community would have to have some way to define an able group of people, perhaps through certification or other demonstrable knowledge of the TEI tag set. Some participants were in favor of a certification process, while others suggested soliciting individual volunteers on a case-by-case basis. The benefit of a certification process is that it would be possible for evaluators to operate both as independent consultants as well as in conjunction with the TEI community. However, defining what constitutes demonstrable knowledge of the TEI tags is a difficult (if not impossible) task, since the organization is mostly an informal one and there is a legitimate potential for a conflict of interest. It is also unclear from the study results *who* should be the party to certify evaluators. Another difficulty in creating a list of approved reviewers and implementing a certification process lies in the funding and infrastructure of the current organization. The overhead alone could be a serious detraction to the certification process. However, the results of this study suggest that there is sufficient interest to continue the exploring how these resources would work in practice. Thus, further discussion of these points within the TEI community would be necessary before any kind of formal certification could proceed.

Secondly, the development of such a list would require the TEI to think further about the logistics of the reviews. The TEI-J would need to consider how to solicit members from the community to perform evaluations of projects, and to determine what benefits the evaluators would receive. That is, the TEI editorial board would need to decide if this activity would be considered professional service and rewarded in terms of publication of the review material. Several participants argued that publication in a peer-reviewed journal was a suitable reward for services, while others suggested that there should be a financial component

supported in the form of a cost-share. Regardless, it is only after determining how to define the body of evaluators and what they could expect from the TEI that a list of capable reviewers could be made available.

The list of reviewers would be made available as a resource both to the TEI and to the wider community. Capable reviewers would, as one participant said, "be able to hang out a shingle" and provide counsel and formative assistance for projects seeking this kind of guidance. It remains to be seen whether these individuals would be compensated by the project that they are assisting or whether this would also be considered a professional service. The list of reviewers would also be made available to other groups performing assessments, such as NINES or institutional tenure and promotion committees. These groups could then work with TEI reviewers to evaluate the text encoding of a project. Compensation would need to be determined in collaboration between the TEI community and the body that they were consulting for, though it is possible that the reviews could be published in the TEI-J following the completion of the review process.

In addition to these suggestions for the TEI, there are a number of questions that still need to be addressed by the general digital humanities community. First and foremost is the need for a clear understanding about the different kinds of evaluation and the standards currently applied in these cases. For the larger community, it is also important to consider how evaluation criteria are defined and applied, and to determine the hierarchy of value for these criteria. Potential next steps in efforts to address these questions may include a survey of digital projects reviews in order to determine what criteria is used in evaluation, how this criteria is applied, and what was the purpose of the review.

Something that could potentially be enlightening to the community and to this topic of evaluation is to consider perspectives outside of the groups that usually define evaluation criteria for the humanities. One particularly valuable resource in this endeavor would be further conversations with the groups that are also working with digital projects but are not a part of the traditional humanities, such as

computer or information science. People from these communities may have a different understanding of tools and be able to offer suggestions about how to evaluate digital scholarship. One participant also made the suggestion that future work might consider interviewing the larger international community, especially those who are operating in countries with nascent digital humanities initiatives. These individuals may face different challenges in evaluating digital projects and approach the same perennial issues from a new perspective. Therefore, addressing the topic of evaluation with underrepresented members of the digital humanities community may potentially add further insight into how digital projects are currently reviewed and possible future development of evaluative criteria.

It would be necessary to consider more about the goals of project evaluations in relation to discipline specific evaluation criteria. While NINES and its sister groups widely praised, some participants noted that NINES was predominantly based on an time period (19th century English and American studies), not necessarily any single discipline (ie, history, linguistics, etc). They argued that a review by a specific community was also necessary in order to address any unique needs of that community. For example, just as NINES evaluates work on 19th century topics, there are specific disciplinary bodies geared toward the evaluation priorities of a single community. The argument was that scholars in specific disciplines look for different features in a project, and each scholar brings a different set of criteria to bear on the project that may or may not reflect the priorities of users outside of this field. For example, one of the participants came from a background in Classics, which traditionally places a higher scholarly value on editing texts. This participant explained that digital projects created and published in this field had specific evaluation criteria that were not necessarily included in guidelines that are targeted primarily towards English and American literature. Another participant, a linguist, agreed and suggested that there were different aspects of a digital project that were evaluated if the project was assessed by another linguist, as opposed to the criteria applied by a historian. In order to

understand how discipline specific review would add to or change the evaluation criteria, it would be necessary to understand more about what evaluation criteria are currently applied and how scholarship is assessed in these contexts. However, in this study participants were not solicited based on a specific discipline within the Humanities, so future work would be necessary to consider what aspects of evaluation criteria are unique and which are shared between disciplines and how evaluative bodies should address these factors.

Determining what to evaluate and how to do so are perennially relevant questions that reflect the values and priorities of scholarship. Digital projects are at an exciting point in their development, and as they continue to challenge traditional evaluation criteria, it will become more important than ever to develop a better understanding of what we value and why we value it. This study has attempted to outline some of these concerns and advocate for a better understanding of the nuances involved in evaluating digital projects.

APPENDIX 1: Survey CFP

11/5/12

Good afternoon,

I am a MSIS candidate at the University of Texas at Austin, School of Information. My faculty sponsor (Dr. Tanya Clement) and I are working on a project that focuses on evaluation methods and standards for digital scholarly projects, with a particular emphasis on those that use the Text Encoding Initiative standard. We would like to gather responses from the TEI community (as well as the general Digital Humanities community) in order to understand more about how digital scholarly projects are currently evaluated. To this end, I have put together a survey and would be grateful if you would take the time to answer a few questions: https://qtrial.qualtrics.com/SE/?SID=SV_b89IXl4msVon4u9

The survey should take no more than 20 minutes of your time. If you take the survey and would be open to discussing this topic further, please contact me at the following email address and we can arrange a virtual interview via Skype at your convenience: SPfannenschmidt@utexas.edu.

Thank you in advance for your assistance with this project; we value your insights.

Best,

Sarah Pfannenschmidt, MSIS candidate, University of Texas at Austin School of Information
Dr. Tanya Clement, Assistant Professor, University of Texas at Austin School of Information

APPENDIX 2: CFP to TEI Conference attendees

Dear [Participant],

I am a MSIS candidate at the University of Texas at Austin, School of Information. My faculty sponsor (Dr. Tanya Clement) and I are working on a project that focuses on on evaluation methods and standards for digital scholarly projects, with a particular emphasis on those that use the Text Encoding Initiative standard. There are two parts to this study: a survey and an interview.

Survey: You can take the survey at any time at https://qtrial.qualtrics.com/SE/SID=SV_b89IXl4msVon4u9

Interview: I will be attending the TEI conference in College Station on Friday, November 9th, 2012 and would like to schedule a short interview (approx. 15 min) with interested participants. If you would be willing to assist us, please either contact me prior to the conference at the email address below to schedule a time or st op by the table at your convenience on Friday. If you are not available for an interview but would still like to provide feedback, send me (Sarah) an email (SPfannenschmidt@utexas.edu) and I can schedule a virtual interview via Skype for a time that better suits your schedule.

Thanks very much for your time, and I look forward to speaking with you.

Regards,
Sarah Pfannenschmidt, MSIS candidate, University of Texas at Austin School of Information
Dr. Tanya Clement, Assistant Professor, University of Texas at Austin School of Information

--
TEI 2012
http://www.tei-c.org/conftool/

APPENDIX 3: Survey

My Report

Last Modified: 03/25/2013

1. Please list up to FIVE digital critical projects that you think exemplify scholarly work.

#	Answer	Bar	Response	%
1	Project 1:		0	0%
2	Project 2:		0	0%
3	Project 3:		0	0%
4	Project 4:		0	0%
5	Project 5		0	0%

Project 1:	Project 2:	Project 3:	Project 4:	Project 5

Statistic	Value
Min Value	-
Max Value	-
Total Responses	0

2. Of the projects you have listed, please identify the area(s) of scholarship to which each belongs.

#	Answer	Bar	Response	%
1	Project 1:		0	0%
2	Project 2:		0	0%
3	Project 3:		0	0%
4	Project 4:		0	0%
5	Project 5:		0	0%

Project 1:	Project 2:	Project 3:	Project 4:	Project 5:

Statistic	Value
Min Value	-
Max Value	-
Total Responses	0

3. In your opinion, what distinguishes the projects you have listed above as 'scholarly'?

Text Response

Statistic	Value
Total Responses	0

4. In general, were there any limitations (technical or content-based) that affected the scholarly merit of the projects that you listed in question 1? Please identify the nature of the limitation(s) and why you consider them to be limitations.

Text Response	
Statistic	**Value**
Total Responses	0

5. In which of the following situations have you used a digital project?

#	Answer	Bar	Response	%
1	For personal reference		0	0%
2	For scholarly pursuit (e.g. research for the publication of a paper, etc.)		0	0%
3	For education (e.g. class resource or preparation for teaching, etc.)		0	0%
4	For comparative analysis (e.g. considering design options for another project, etc.)		0	0%
5	Other use (please specify)		0	0%

Other use (please specify)	
Statistic	**Value**
Min Value	-
Max Value	-
Total Responses	0

6. Are you familiar with any current standards used by the TEI community to evaluate the scholarship of digital critical projects?

#	Answer	Bar	Response	%
1	Yes, I am aware of evaluation standards used by the community		0	0%
2	I think that there are evaluation standards but do not know what they are (if 'maybe', please skip to question 13)		0	0%
3	No, I am not aware of any evaluation standards used by the community (if 'no', please skip to question 13)		0	0%
	Total		0	

Statistic	Value
Min Value	-
Max Value	-
Mean	0.00
Variance	0.00
Standard Deviation	0.00
Total Responses	0

7. If you answered 'yes' to question 6, please describe what these evaluation standards are in your own words.

Text Response	
Statistic	**Value**
Total Responses	0

8. Where were you introduced to these evaluation standards?

#	Answer	Bar	Response	%
1	Single source: standards are outlined as part of a single project (please specify)		0	0%
2	No single source: standards articulated in part across multiple projects (please specify)		0	0%
3	No single source: standards are part of shared, but not articulated, knowledge within the community		0	0%
4	Other (please specify)		0	0%
	Total		0	

Single source: standards are outlined as part of a single project (please specify)	No single source: standards articulated in part across multiple projects (please specify)	Other (please specify)

Statistic	Value
Min Value	-
Max Value	-
Mean	0.00
Variance	0.00
Standard Deviation	0.00
Total Responses	0

9. Do you believe these evaluation standards to be widely accepted?

#	Answer	Bar	Response	%
1	Definitely yes		0	0%
2	Probably yes		0	0%
3	Maybe		0	0%
4	Probably not		0	0%
5	Definitely not		0	0%
	Total		0	

Statistic	Value
Min Value	-
Max Value	-
Mean	0.00
Variance	0.00
Standard Deviation	0.00
Total Responses	0

10. In your opinion, why/why not are these evaluation standards widely accepted?

Text Response

Statistic	Value
Total Responses	0

11. In your opinion, are there limitations to the evaluation standards used by the community or the way in which they are currently applied?

#	Answer	Bar	Response	%
1	Definitely yes		0	0%
2	Probably yes		0	0%
3	Maybe		0	0%
4	Probably not		0	0%
5	Definitely not		0	0%
	Total		0	

Statistic	Value
Min Value	-
Max Value	-
Mean	0.00
Variance	0.00
Standard Deviation	0.00
Total Responses	0

12. If you answered yes to question 10, what do you feel are the limitations of these standards?

Text Response

Statistic	Value
Total Responses	0

13. What is the role of the TEI council in the creation of evaluation standards for identifying 'scholarly' digital projects?

#	Answer	Bar	Response	%
1	The council should lead the community in the creation of evaluation standards		0	0%
2	The council should encourage others to develop the standards but should not seek to actively influence the standards		0	0%
3	The council should not be involved in the creation of evaluation standards (if 'no', please skip to question 15)		0	0%
4	Other (please specify)		0	0%
	Total		0	

Other (please specify)

Statistic	Value
Min Value	-
Max Value	-
Mean	0.00
Variance	0.00
Standard Deviation	0.00
Total Responses	0

14. If you think that the TEI council should be part of the conversation of evaluation standards development, what form should these contributions take?

#	Answer	Bar	Response	%
1	General reports on evaluation standards issued to the community on a regular basis		0	0%
2	Implement a temporary committee or task force to appraise the issue and develop a system of evaluation		0	0%
3	Implement a permanent committee or task force to develop and continually evaluate the standards by the community		0	0%
4	Other (please specify)		0	0%
	Total		0	

Other (please specify)	
Statistic	**Value**
Min Value	-
Max Value	-
Mean	0.00
Variance	0.00
Standard Deviation	0.00
Total Responses	0

15. If you have been a part of the creation of a digital project, what was/were the goal(s) of the project and your role in it?

Text Response	
Statistic	**Value**
Total Responses	0

16. Was the project evaluated or reviewed in any way before or after completion?

#	Answer	Bar	Response	%
1	Yes		0	0%
2	No		0	0%
3	Do not know if the project was evaluated or reviewed		0	0%

Statistic	Value
Min Value	-
Max Value	-
Total Responses	0

17. Who PRIMARILY evaluated the digital project?

#	Answer	Bar	Response	%
1	Internal evaluation performed by peers not associated with Digital Humanities within affiliated institution(s)		0	0%
2	External evaluation performed by peers not associated with Digital Humanities from non-affiliated institution(s)		0	0%
3	Internal evaluation performed by peers associated with Digital Humanities within affiliated institution(s)		0	0%
4	Evaluation performed by a Digital Humanities peer-review organization (e.g. NINES, 18thConnect, etc.)		0	0%
5	Evaluation performed by grant funding body		0	0%
6	Other (please specify)		0	0%
	Total		0	

Other (please specify)

Statistic	Value
Min Value	-
Max Value	-
Mean	0.00
Variance	0.00
Standard Deviation	0.00
Total Responses	0

18. What standards were applied in the evaluation of this project?

#	Answer	Bar	Response	%
1	Individual standards from participating institution(s) outlined prior to the commencement of the project		0	0%
2	Standards taken from governing body (e.g. MLA or equivalent professional organization)		0	0%
3	Standards taken from guidelines given by a grant funding body (e.g. NEH or equivalent)		0	0%
4	Other (please specify)		0	0%

Other (please specify)

Statistic	Value
Min Value	-
Max Value	-
Total Responses	0

19. In your opinion, what is the relationship between the content of a digital critical project and the tools used to create it in determining the project's overall scholarly value?

#	Answer	Bar	Response	%
1	Evaluation is equally based on content and tools		0	0%
2	Evaluation is based predominantly on content		0	0%
3	Evaluation is based only on tools		0	0%
4	Evaluation not based on content or tools		0	0%
5	Other (please specify)		0	0%
	Total		0	

Other (please specify)

Statistic	Value
Min Value	-
Max Value	-
Mean	0.00
Variance	0.00
Standard Deviation	0.00
Total Responses	0

20. What aspects of a project should be evaluated when determining the project's scholarly value? Please list them in order of importance, with 100 representing 'very important' and 0 representing 'no importance in evaluation'.

#	Answer	Min Value	Max Value	Average Value	Standard Deviation	Responses
1	Interface design	0.00	0.00	0.00	0.00	-
2	Innovative use of tools	0.00	0.00	0.00	0.00	-
3	Rigor of encoding	0.00	0.00	0.00	0.00	-
4	Impact on the scholarly community	0.00	0.00	0.00	0.00	-
5	Clearly defined and implemented editorial practices	0.00	0.00	0.00	0.00	-
6	Contribution to existing body of knowledge	0.00	0.00	0.00	0.00	-
7	Usability	0.00	0.00	0.00	0.00	-
8	Interoperability	0.00	0.00	0.00	0.00	-
9	Sustainability	0.00	0.00	0.00	0.00	-
10	Accessibility	0.00	0.00	0.00	0.00	-
11	Other criteria [please specify].	0.00	0.00	0.00	0.00	-

Other criteria [please specify].

APPENDIX 4: Interview Consent Form

Informed Consent for Participation in the Research Project, "Evaluation Standards for Digital Scholarship"

Introduction

The purpose of this form is to provide you information that may affect your decision as to whether or not to participate in this research study by being interviewed. The person conducting the interview [Sarah Pfannenschmidt] will answer any of your questions. Read the information below and ask any questions you might have before deciding whether or not to take part. If you decide to be interviewed, this form will be used to record your consent.

Purpose of the Project

You have been asked to participate in a research project about the creation, use, and evaluation of scholarly digital projects. The purpose of this study is to understand how current scholars in the field of the humanities and/or computing use and evaluate digital critical scholarship.

What will you to be asked to do?

If you agree to participate in this study, you are agreeing to be interviewed. The interview is expected to take between 15 and 30 minutes of your time. You will be asked a series of questions, any of which you may refuse to answer.

NOTE: If you choose to participate in this study, your responses may be audio recorded. **The recording will in no way be shared with anyone not affiliated with this project and will be available only to the researcher for reference purposes. The recording will be kept for up to 72 hours after your interview, at which point it will be transcribed and the audio file will be permanently overwritten. Personal information such as name, institutional affiliation, etc. mentioned during the interview will not be used in the discussion of the project results.** The data from your participation may be used for future research. Any non-private information may be made available for research purposes not detailed within this consent form at the discretion of the principle investigator [Sarah Pfannenschmidt] or the faculty sponsor [Dr. Tanya Clement].

What are the possible risks and benefits of this project?

There are no foreseeable risks to participating in this project. While you will receive no direct benefit or any type of payment from participating, be aware that your participation contributes to a better understanding of current work and evaluation standards of scholarship in the field of humanities and computing.

Do you have to participate?

No, your participation is voluntary. You may decide not to participate at all or, if you start the interview, you may withdraw at any time.

If you are willing to participate, please sign, date, and return the second copy of the document to the interviewer prior to the interview. If the interview is undertaken in person, you will be asked to sign and date two forms, one of which will be given to you for your records. In the event of a long-distance interview, you will receive a copy of this form.

Whom to contact with questions about the project?

Prior, during or after your participation you can contact the researcher, Sarah Pfannenschmidt, at 000-000-0000, or send an email to SPfannenschmidt@utexas.edu.

Signature

You have been informed about this study's purpose, procedures, possible benefits and risks, and you have received a copy of this form. You have been given the opportunity to ask questions before you sign, and you have been told that you can ask other questions at any time. You voluntarily agree to participate in this project. By signing this form, you are not waiving any of your legal rights.

_____
Printed Name

_____ _____
Signature Date

As a representative of this study, I have explained the purpose, procedures, benefits, and the risks involved in this research project.

_____
Print Name of Person obtaining consent

_____ _____
Signature of Person obtaining consent Date

WORKS CITED

1. Thornton, John K. "Disseminating History in the Digital Age." American Historical Association, *Perspectives on History*. January 2011.
2. Rieger, Oya. "Framing digital humanities: The role of new media in humanities scholarship." *First Monday* 15.10 (October 2010).
3. Berry, Dave. "The Computational Turn: Thinking about the Digital Humanities." *Culture Machine* 11 (2011): 1-22.
4. Welshons, Marlo. "What is Digital Scholarship?" Connexions. 11 Dec. 2006 http://cnx.org/content/m14163/1.1/. Accessed April 19th, 2013.
5. Mandell, Laura. "Promotion for Tenure and Scholarship" in *Journal Of Digital Humanities* 1.4 (Fall 2012): 46-53.
6. Kirschenbaum, Matthew G. "What is Digital Humanities and What's it Doing in English Departments?" *ADE Bulletin* 150 (2010): 55-61.
7. Hayles, N. Katherine. "How We Think: Transforming Power and Digital Technologies" in *Understanding Digital Humanities*, ed. David M. Berry. Palgrave MacMillan, (2012): 42-66.
8. Unsworth, John. "What is Humanities Computing, and What is Not?" in *Jahrbuch für Computerphilologie 4*, Georg Braungart, Karl Eibl & Fotis Jannidis, eds. Paderborn: mentis 2002.
9. Rockwell, Geoffrey. "How do you define Digital Humanities?" from the Day of DH 2012. University of Alberta, Canada. http://dayofdh2012.artsrn.ualberta.ca/dh/. Accessed March 19th, 2013.
10. Crane, Gregory, Bamman, David, Cerrato, Lisa, Jones, Alison, Mimno, David, Packel, Adrian, Sculley, David, and Gabriel Weaver. "Beyond Digital Incunabula: Modeling the Next Generation of Digital Libraries." *Proceedings of the ECDL* (2006): 353-366.
11. Kets, Annemarie. "Texts Worth Editing: Polyperspectival Corpora of Letters." *Variants; The Journal of the European Society for Textual Scholarship* 10 (2013): 93-103.
12. Palmer, Carole, Zavalina, Oksana, and Megan Mustafoff. "Trends in Metadata Practices: A Longitudinal Study of Collection Federation." *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries* (2007): 386-395.
13. Zorich, Diane M. "A Survey of Digital Humanities Centers in the United States." Washington, DC: Council on Library and Information Resources (2008).
14. Schriebman, Susan, Laura Mandell, and Stephen Olsen. "Evaluating Digital Scholarship: Introduction." *Profession* (2011): 123-35.
15. Cohen, Daniel J., and Joan Fragaszy Troyano. "Closing the Evaluation Gap." *Journal Of Digital Humanities* 1.4 (Fall 2012): i-ii.
16. NINES. "Evaluating Digital Scholarship." NINES/NEH Summer Institutes: 2011-2012. http://institutes.nines.org/. Accessed December 29th, 2012.

17. Purdy, James P., and Joyce R. Walker. "Valuing Digital Scholarship: Exploring the Changing Realities of Intellectual Work." *Profession* (2010): 177-95.
18. Tanselle, G. Thomas. "Foreword" in *Electronic Textual Editing*, eds. Unsworth, John, Lou Burnard, and Katherine O'Brien O'Keeffe. New York: Modern Language Association of America. 2006, 1-6.
19. McGann, Jerome. *Radiant Textuality: Literature after the World Wide Web.* New York: Palgrave, 2002.
20. Mandell, Laura. "Special Issue: 'Scholarly Editing in the Twenty-First Century'-A Conclusion." *Literature Compass* 7.2 (2010): 120-33.
21. Curran, Stuart. "Different Demands, Different Priorities: Electronic and Print Editions." *Literature Compass* 7.2 (2010): 82-8.
22. Presner, Todd. "How to Evaluate Digital Scholarship" in *Journal of Digital Humanities* 1.4 (Fall 2012): 35-39.
23. Anderson, Steve, and Tara McPherson. "Engaging Digital Scholarship: Thoughts on Evaluating Multimedia Scholarship." *Profession* (2011): 136-51.
24. Smith, Martha Nell. "The Human Touch Software of the Highest Order: Revisiting Editing as Interpretation." *Textual Cultures* 2.1 (Spring 2007): 1-15.
25. Cavanagh, Sheila. "Living in a Digital World: Rethinking Peer Review, Collaboration, and Open Access." *Journal of Digital Humanities* 1.4 (Fall 2012): 3-16.
26. Nowviskie, Bethany. "Where Credit is due: Preconditions for the Evaluation of Collaborative Digital Scholarship." *Profession* (2011): 169-81.
27. Burgess, Helen J., and Jeanne Hamming. "New Media in the Academy: Labor and the Production of Knowledge in Scholarly Multimedia." *Digital Humanities Quarterly* 5.3 (2011).
28. Flanders, Julia. "The Productive Unease of 21-st Century Digital Scholarship." *Digital Humanities Quarterly* 3.3 (2009).
29. Unsworth, John. "Documenting the Reinvention of Text: The Importance of Failure." *Journal of Electronic Publishing* 3.2 (December 1997).
30. MLA Task Force on Evaluating Scholarship for Tenure and Promotion. "Report of the MLA Task Force on Evaluating Scholarship for Tenure and Promotion." *Profession* (2007): 9-71.
31. The Working Group on Evaluating Public History Scholarship. *Tenure, Promotion, and the Publicly Engaged Academic Historian*. OAH Executive Board, NCPH, and AHA Council, 2010. Accessed December 15th, 2012.
32. Institute for the Study of the Ancient World, New York University. "Assessing Research at ISAW." http://isaw.nyu.edu/research/AssessingresearchatISAW.pdf (2012). Accessed December 18th, 2012.
33. NEH Digital Humanities Startup Grants Guidelines http://www.neh.gov/grants/odh/digital-humanities-start-grants. Accessed October 18th, 2012.

34. Blair, Kristine L., Gail E. Hawisher, and Cynthia L. Selfe. "The Electronic Landscape of Journal Editing: *Computers and Composition* as a Scholarly Collective." *Profession* (2009): 160-7.
35. NINES. "General Guidelines and Peer Review for NINES Content." http://www.nines.org/about/wp-content/uploads/2011/12/9s-guidelines.doc. 2012. Accessed November 1st, 2012.
36. Gabler, Hans Walter. "Theorizing the Digital Scholarly Edition." *Literature Compass* 7.2 (2010): 43-56.
37. Price, Kenneth. "Electronic Scholarly Editions." *A Companion to Digital Literary Studies.* Eds. Ray Siemans and Susan Schriebman. Malden, MA: Blackwell Publishing, 2007.
38. Robinson, Peter. "Towards a Theory of Digital Editions." *Variants; The Journal of the European Society for Textual Scholarship* 10 (2013): 105-131.
39. Pierazzo, Elena. "A rationale of digital documentary editions." *Literary and Linguistic Computing* 26.4 (2011): 463-477.
40. Clement, Tanya. "Knowledge Representation and Digital Scholarly Editions in Theory and Practice." *Journal of the Text Encoding Initiative*. 1 (2011).
41. Shillinsburg, Peter L. "Polymorphic, Polysemic, Protean, Reliable, Electronic Texts." *Palimpsest: Editorial Theory in the Humanities*. Eds. George Bornstein and Ralph G. Williams. Ann Arbor, Michigan: University of Michigan Press, 1993. 29-43.
42. Cross, Jeanne G. "Reviewing Digital Scholarship: The Need for Discipline-Based Peer Review." *Journal of Web Librarianship* 2.4 (2008): 549-66.
43. Greetham, D. C. "The Resistance to Philology." Ed. D.C. Greetham. Ann Arbor, Michigan: University of Michigan Press, 1997. 9-24.
44. TEI Website. "Projects Using the TEI." http://www.tei-c.org/Activities/Projects/. Accessed April 9th, 2013.
45. Cummings, James. "The Text Encoding Initiative and the Study of Literature." *A Companion to Digital Literary Studies*. Eds. Ray Siemens and Susan Schriebman. Oxford: Blackwell Publishing, 2007. 451-476.
46. Renear, Alan. "Text Encoding." *A Companion to Digital Humanities*. Eds. Susan Schriebman, Ray Siemens, and John Unsworth. Oxford: Blackwell Publishing, 2004. 218-239.
47. Sperberg-McQueen, C.M. *Textual Criticism and the Text Encoding Initiative*. Annual Convention of the Modern Language Association. http://www.tei-c.org/Vault/XX/mla94.html. Accessed April 13th, 2013.
48. Burghart, Marjorie, and Malte Rehbein. "The Present and Future of the TEI Community for Manuscript Encoding." *Journal of the Text Encoding Initiative*. Issue 2 (February 2012).
49. Nowviskie, Bethany, ed. *"#Alt-Academy"*, Media Commons, May 2012. http://mediacommons.futureofthebook.org/alt-ac/welcome. Accessed March 23rd, 2013.

50. Fitzpatrick, Kathleen. "Peer Review, Judgment, and Reading." *Profession* (2011): 196-201.
51. Jannidis, Fotis. "TEI in a Crystal Ball." *Literary and Linguistic Computing* 24.3 (2009): 253-265.
52. TEI homepage, "TEI: The Text Encoding Initiative." http://www.tei-c.org/index.xml. Accessed March 17th, 2013.
53. NINES website, 'What is Nines?' www.nines.org/about/what-is-nines/. Accessed March 13th, 2013.
54. Lavagnino, John. "When Not To Use TEI." *Electronic Textual Editing*, eds. Unsworth, John, Lou Burnard, and Katherine O'Brien O'Keeffe. New York: Modern Language Association of America. 2006. 334-348.
55. Fiormonte, Domenico, Valentina Miratiradonna, and Desmond Schmidt. "Digital Encoding as a Hermeneutic and Semiotic Act: The Case of Valerio Magrelli." *Digital Humanities Quarterly* 4.1 (2010).
56. Rockwell, Geoffrey. "On the Evaluation of Digital Media as Scholarship." *Profession* (2011): 152-67.